

Search Notes

L Number	Hits	Search Text	DB	Time stamp
1	31	hiwi or piwi	USPAT; US-PGPUB; EPO; JPO; DERWENT	2003/12/13 18:18
2	19	(hiwi or piwi) and human	USPAT; US-PGPUB; EPO; JPO; DERWENT	2003/12/13 18:18
3	8	(hiwi or piwi) same human	USPAT; US-PGPUB; EPO; JPO; DERWENT	2003/12/13 18:18

Set	Items	Description
S1	220	HIWI OR PIWI
S2	95	RD (unique items)

XX

PR 04-DEC-1998; 98US-0110901.
XX (UYDU-) UNIV DUKE.
XX Lin H;
XX WPI: 2000-412085/35.
XX N-PSDB; AAA07588.
XX Piwi family nucleic acids, polypeptides, and antibodies, useful in gene
XX therapy of diseases such as cancer and in various research and
XX diagnostic applications -
XX Claim 4; Page 189-194; 201pp; English.
XX This sequence represents the human piwi family protein, designated
XX hiwi. The piwi family nucleic acids and polypeptides are used in gene
XX therapy of diseases such as cancer and also in various research and
XX diagnostic applications. The sequences can also be used to treat
XX tissue dystrophy, anaemia, immunodeficiency, and male infertility.
XX
SQ Sequence 861 AA;
Query Match 99.0%; Score 4061; DB 21; Length 861;
Best Local Similarity 99.5%; Pred. No. 0;
Matches 768; Conservative 0; Mismatches 4; Indels 0; Gaps 0;
QY 4 GVNTRQNLHVKESKTGSSGIIVRLSTNHFRLTSRPQWALYQYHIDYNPLMEARLRSL 63
Db 90 GVNTRQNLHVKESKTGSSGIIVRLSTNHFRLTSRPQWALYQYHIDYNPLMEARLRSL 149
QY 64 LFOHEDLIGKCHAFDGTILFLPKRLQKQVTEVFSKTRNGEDVRITITLTNELPPTSPTCL 123
Db 150 LFOHEDLIGKCHAFDGTILFLPKRLQKQVTEVFSKTRNGEDVRITITLTNELPPTSPTCL 209
QY 124 QFYNIIFRLKIMLQIQRNYNPNDPIDPSHRLVWPGETTSIIQYENSIMLCTDV 183
Db 210 QFYNIIFRLKIMLQIQRNYNPNDPIDPSHRLVWPGETTSIIQYENSIMLCTDV 269
QY 184 SHKVLRSVTLDFMNFYHQTTEHKFQGVSKELIGLVLTIKYNNKTYRVDIDWDQNP 243
Db 270 SHKVLRSVTLDFMNFYHQTTEHKFQGVSKELIGLVLTIKYNNKTYRVDIDWDQNP 329
QY 244 STFKKADGSEVSFLYRKQYNOEITDLKQPLVSPQRRRPGGTLPGLMILPELCYL 303
Db 330 STFKKADGSGVFLYRKQYNOEITDLKQPLVSPQRRRPGGTLPGLMILPELCYL 389
QY 304 TGLTDKMNDFNMKDLAVHTLTPQQRQREYGRLLIDYHKNDNVORELRDWGLSFDNL 363
Db 390 TGLTDKMNDFNMKDLAVHTLTPQQRQREYGRLLIDYHKNDNVORELRDWGLSFDNL 449
QY 364 LSPSGRIILQTEKHGGKTFDYNPQFADWSKETRGAPLISVKPLDNWLLIYTRRYEAA 423
Db 450 LSPSGRIILQTEKHGGKTFDYNPQFADWSKETRGAPLISVKPLDNWLLIYTRRYEAA 509
QY 424 SLIQNLKVTTPANGQMKAMIEVDDETEALVRLVLOKVTADTQIVVCLSSNKKDYD 483
Db 510 SLIQNLKVTTPANGQMKAMIEVDDETEALVRLVLOKVTADTQIVVCLSSNKKDYD 569
QY 484 AIKKYLCTDPTSCQVAVTGLKQOTVMAIATKIALQNMCKMGELWRVDIPLKLVMI 543
Db 570 AIKKYLCTDPTSCQVAVTGLKQOTVMAIATKIALQNMCKMGELWRVDIPLKLVMI 629
QY 544 GIDCHDMTAGRRSIAGFVASINEGTRWFSRIFQDRGOELVDGLKVLQALRAMNSC 603
Db 630 GIDCHDMTAGRRSIAGFVASINEGTRWFSRIFQDRGOELVDGLKVLQALRAMNSC 689
QY 604 NEYMSRIIVYRDGVDGQOLKTLVNYEVPFLDCLKSI GRGNPRLTVIVVKRVTNTRFF 663
Db 690 NEYMSRIIVYRDGVDGQOLKTLVNYEVPFLDCLKSI GRGNPRLTVIVVKRVTNTRFF 749
QY 664 AQSGRLQNLPGTVIDVETRPENYDFFIIVSQAVRSGSVSPTHYVNDYNSGLKPDHIQ 723

Db 750 AQSGRLQNLPGTVIDVETRPENYDFFIIVSQAVRSGSVSPTHYVNDYNSGLKPDHIQ 809
QY 724 RLTYKLCHIIYNNWPGVIRVPAPCOYAHKLAFLVGCQSIHREPNLSLSNRLYYL 775
Db 810 RLTYKLCHIIYNNWPGVIRVPAPCOYAHKLAFLVGCQSIHREPNLSLSNRLYYL 861
RESULT 2
AAAY90234
ID AAAY90234 standard; Protein; 862 AA.
XX AC AAAY90234;
XX DT 29-AUG-2000 (first entry)
XX Mouse piwi protein, designated miwi.
XX Piwi family protein; piwi; miwi; hiwi; gene therapy; tissue dystrophy;
XX anaemia; immunodeficiency; male infertility; mouse; ds.
XX Mus sp.
XX Key Location/Qualifiers
FH Misc-difference 90 /label= Leu, Ile
FT /note= "encoded by NTC"
FT Misc-difference 216 /note= "unspecified amino acid, encoded by NTC"
FT Misc-difference 383 /label= Leu, Ile
FT /note= "encoded by NTC"
FT Misc-difference 816 /label= Leu, Ile
FT /note= "encoded by NTC"
XX WO2000032039-A1.
XX 08-JUN-2000.
XX 03-DEC-1999; 99WO-US28764.
XX 04-DEC-1998; 98US-0110901.
XX (UYDU-) UNIV DUKE.
XX Lin H;
XX WPI: 2000-412085/35.
XX N-PSDB; AAA07587.
XX Piwi family nucleic acids, polypeptides, and antibodies, useful in gene
XX therapy of diseases such as cancer and in various research and
XX diagnostic applications -
XX Claim 4; Page 180-185; 201pp; English.
XX This sequence represents the mouse piwi family protein, designated
XX miwi. The piwi family nucleic acids and polypeptides are used in gene
XX therapy of diseases such as cancer and also in various research and
XX diagnostic applications. The sequences can also be used to treat
XX tissue dystrophy, anaemia, immunodeficiency, and male infertility.
XX
SQ Sequence 862 AA;
Query Match 96.5%; Score 3957; DB 21; Length 862;
Best Local Similarity 95.9%; Pred. No. 0;
Matches 740; Conservative 19; Mismatches 13; Indels 0; Gaps 0;
QY 4 GVNTRQNLHVKESKTGSSGIIVRLSTNHFRLTSRPQWALYQYHIDYNPLMEARLRSL 63
Db 91 GVNTRQNLHVKESKTGSSGIIVRLSTNHFRLTSRPQWALYQYHIDYNPLMEARLRSL 150
QY 64 LFOHEDLIGKCHAFDGTILFLPKRLQKQVTEVFSKTRNGEDVRITITLTNELPPTSPTCL 123


```
Db 151 LFOHEDLIQRCHAFDGTILFLKRLQHKVTEFSQTRNGEHRVITITLTNELPPTSPCL 210
Qy 124 QFYNIIFRLLKIMNLOQIGRNYNPNPDIDIPSHRLVWPFGFTTSILQYENSIMLCYDV 183
Db 211 QFYNIIFRLLKIMNLOQIGRNYNPNPDIDIPSHRLVWPFGFTTSILQYENNIMLCYDV 270
Qy 184 SHKVLRSVTLDFMFNFYHOTEHFFQEQVSKELIGLVLTLYNNKTYRVDIDWDQPK 243
Db 271 SHKVLRSVTLDFMFNLVQOTEHFFQEQVSKELIGLVLTLYNNKTYRVDIDWDQPK 330
Qy 244 STFKKADGSEVSFLYRKQYNQEIITDLKQPVLSQPKRRRPGGTLPFAPMLIPELCYL 303
Db 331 STFKKADGSEVSFLYRKQYNQEIITDLKQPVLSQPKRRRPGGTLPFAPMXIFELCYL 390
Qy 304 TGLTDXMENDFNKDLAVHTLTPEQOREVGRLLDIYHKNDNVQRELWDGLSFDNSL 363
Db 391 TGLTDXMENDFNKDLAVHTLTPEQOREVGRLLDIYHKNDNVQRELWDGLSFDNSL 450
Qy 364 LSFSGRIILQTEKHOGGKTFDYNPOFADWSKETRGAPLISVKPLDNWLLIYTRRYEAA 423
Db 451 LSFSGRIILQSEKIHOGGKTFDYNPOFADWSKETRGAPLISVKPLDNWLLIYTRRYEAA 510
Qy 424 SLIQNLFKVTPAMGQMKAMIEVDVDRTEAYLRVLQOKVTADTOIVVCLLSNRRKDYD 483
Db 511 SLIQNLFKVTPAMGQMKAMIEVDVDRTEAYLRVLQOKVTSDTQIVVCLLSNRRKDYD 570
Qy 484 ALKKYLCTDCTPSPQCVARTLGKQOTVMAIATKIALQMNCKMGELWRVDIPLKLVIV 543
Db 571 ALKKYLCTDCTPSPQCVARTLGKQOTVMAIATKIALQMNCKMGELWRVDMALAMIV 630
Qy 544 GIDCVHDMTAGRSIAGFVASINEGTRWFSCIFQDRGOELVDGLKVCLOALRAWNSC 603
Db 631 GIDCVHDMTAGRSIAGFVASINEGTRWFSCIFQDRGOELVDGLKVCLOALRAWNSC 690
Qy 604 NEYMPRIIVYRDGVGDGQLKTLVNYEVPQFLDCLKSIGRGYNPRLTVIVVKRYNTRFF 663
Db 691 NEYMPRIIVYRDGVGDGQLKTLVNYEVPQFLDCLKSIGRGYNPRLTVIVVKRYNTRFF 750
Qy 664 AQSGRLQNLPGTVIDVEVTPPEWDFPIVQAVRSGSVSTHNYIYDNSGLKPDHIQ 723
Db 751 AQSGRLQNLPGTVIDVEVTPPEWDFPIVQAVRSGSVSTHNYIYDSSGLKPDHIQ 810
Qy 724 RLTYKLCCHIYNNWPVIRVAPCOYAHKLAFLVGSIHREPNLSNRLYYL 775
Db 811 RLTYKLCCHIYNNWPVIRVAPCOYAHKLAFLVGSIHREPNLSNRLYYL 862

RESULT 3
ID ABP67995 standard; Protein; 523 AA.
XX
AC ABP67995;
XX
DT 13-DEC-2002 (first entry)
XX
DE Human colon cancer related polypeptide SEQ ID NO 2603.
XX
KW Human; colon; cancer; cytostatic; tumour; gene therapy; vaccine.
XX
OS Homo sapiens.
XX
PN WO200258534-A2.
XX
PD 01-AUG-2002.
XX
PF 19-NOV-2001; 2001WO-US43704.
XX
PR 20-NOV-2000; 2000US-252222P.
PR 06-FEB-2001; 2001US-267011P.
PR 28-MAR-2001; 2001US-279670P.
PR 10-JUL-2001; 2001US-304037P.
XX
```

(CORI-) CORIXA CORP.

Stolk JA, Xu J, Chenault RA, Meagher MJ, Secrist H, King GB;

WPI; 2002-608400/65.

N-PSDB; ABV89287.

New isolated tumor colon polynucleotide and polypeptide, useful for the diagnosis, prevention and/or treatment of cancer, in particular colon cancer

Claim 2; SEQ ID NO 2603; 266pp + Sequence Listing; English.

The invention relates to a human colon tumour expressed polynucleotide (I) encoding a polypeptide (II, ABP67991-ABP67996) comprising: (i) any of 2600 fully defined nucleotide sequences (ABV8669-ABV89289); (ii) complements of (i); (iii) at least 20 contiguous residues of (i); (iv) sequences that hybridize to (i), under moderately stringent conditions; (v) sequences having at least 75% or 90% identity to (i); or (vi) degenerate variants of (i). The compositions and methods of the present invention are useful for the diagnosis, prevention and/or treatment of cancer, particularly colon cancer. (i) can be used in gene therapy and (II) and (II) are useful in pharmaceutical compositions such as vaccines. Note: The sequence data for this patent did not form part of the printed specification, but was obtained in electronic format directly from WIPO at ftp.wipo.int/pub/published_pct_sequences.

SQ Sequence 523 AA;

Query Match 67.4%; Score 2764; DB 23; Length 523;

Best Local Similarity 100.0%; Pred. No. 9,7e-261;

Matches 522; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 254 VSFLEYRKYQNOEITDLKQPVLSQPKRRRPGGTLPFAPMLIPELCYLGLTDMRND 313

Db 2 VSFLEYRKYQNOEITDLKQPVLSQPKRRRPGGTLPFAPMLIPELCYLGLTDMRND 61

Qy 314 FNVKMDLAVHRLTPEQOREVGRLLDIYHKNDNVQRELWDGLSFDNSLSFSGRILOT 373

Db 62 FNVKMDLAVHRLTPEQOREVGRLLDIYHKNDNVQRELWDGLSFDNSLSFSGRILOT 121

Qy 374 EKHOGGKTFDYNPOFADWSKETRGAPLISVKPLDNWLLIYTRRYEAAANSIQLFKVT 433

Db 122 EKHOGGKTFDYNPOFADWSKETRGAPLISVKPLDNWLLIYTRRYEAAANSIQLFKVT 181

Qy 434 PAMGQMKAMIEVDVDRTEAYLRVLQOKVTADTOIVVCLLSNRRKDYDAIKKYLCTDC 493

Db 182 PAMGQMKAMIEVDVDRTEAYLRVLQOKVTADTOIVVCLLSNRRKDYDAIKKYLCTDC 241

Qy 494 PTPSQVVARLTGKQOTVMAIATKIALQMNCKMGELWRVDIPLKLVIVGDCYHDMTA 553

Db 242 PTPSQVVARLTGKQOTVMAIATKIALQMNCKMGELWRVDIPLKLVIVGDCYHDMTA 301

Qy 554 GRSIAGFVASINEGTRWFSCIFQDRGOELVDGLKVCLOALRAWNSCNEYMPRIIV 613

Db 302 GRSIAGFVASINEGTRWFSCIFQDRGOELVDGLKVCLOALRAWNSCNEYMPRIIV 361

Qy 614 YRDGVGDGQLKTLVNYEVPQFLDCLKSIGRGYNPRLTVIVVKRYNTRFFAQSGRLQNP 673

Db 362 YRDGVGDGQLKTLVNYEVPQFLDCLKSIGRGYNPRLTVIVVKRYNTRFFAQSGRLQNP 421

Qy 674 LPGTVIDVEVTPPEWDFPIVQAVRSGSVSTHNYIYDNSGLKPDHIQRLTYKLCCHIY 733

Db 422 LPGTVIDVEVTPPEWDFPIVQAVRSGSVSTHNYIYDNSGLKPDHIQRLTYKLCCHIY 481

Qy 734 YNWPVIRVAPCOYAHKLAFLVGSIHREPNLSNRLYYL 775

Db 482 YNWPVIRVAPCOYAHKLAFLVGSIHREPNLSNRLYYL 523

RESULT 4

ABP62102

ID ABP62102 standard; Protein; 866 AA.

XX ABB62102;
 XX AC
 XX DT 26-MAR-2002 (first entry)
 XX DE Drosophila melanogaster polypeptide SEQ ID NO 13098.
 XX KW Drosophila; developmental biology; cell signalling; insecticide;
 XX KW pharmaceutical.
 XX OS Drosophila melanogaster.
 XX PN WO200171042-A2.
 XX PD 27-SEP-2001.
 XX PF 23-MAR-2001; 2001WO-US09231.
 XX PR 23-MAR-2000; 2000US-191637P.
 XX PR 11-JUL-2000; 2000US-0614150.
 XX PA (PEKE) PE CORP NY.
 XX PI Venter JC, Adams M, Li PWD, Myers EW;
 XX PX WPI; 2001-656860/75.
 XX DR N-PSDB; ABL06205.
 XX DT New isolated nucleic acid detection reagent for detecting 1000 or more
 XX PT genes from Drosophila and for elucidating cell signalling and cell-cell
 XX PT interactions -
 XX PS Disclosure; SEQ ID NO 13098; 21pp + Sequence Listing; English.
 XX CC The invention relates to an isolated nucleic acid detection reagent
 XX CC capable of detecting 1000 or more genes from Drosophila. The invention is
 XX CC useful in developmental biology and in elucidating cell signalling and
 XX CC cell-cell interactions in higher eukaryotes for the development of
 XX CC insecticides, therapeutics and pharmaceutical drugs. The invention
 XX CC discloses genomic DNA sequences (ABL16176-ABL30511), expressed DNA
 XX CC sequences (ABL01840-ABL16175) and the encoded proteins
 XX CC (ABB57737-ABB72072).
 XX CC The sequence data for this patent did not form part of the printed
 XX CC specification, but was obtained in electronic format directly from WIPO
 XX CC at ftp.wipo.int/pub/published_pct_sequences.
 XX SQ Sequence 866 AA;
 Query Match 35.8%; Score 1468; DB 22; Length 866;
 Best Local Similarity 39.4%; Pred. No. 1.2e-133;
 Matches 305; Conservative 154; Mismatches 291; Indels 24; Gaps 12;
 QY 17 SKTGSGIIVRLSTNHFRLTSRPOWALYOHIDYNPLMEARRLSALLFQHEDLGKCHA 76
 DB 102 SKGVGVGTHITVQANFKVLRNWTIYQYVDFTDVEATRLRRSFLYEHKGLGG-VI 160
 QY 77 FQGTILFLP--KRLQOK--VTEVFKTRNGEDVRITITLNLPEPTSPCTCLQFYNIIFR 131
 DB 161 FQGTNMFICNQFKAVQDSFVLELVTKSRAGENIEIKIKAVGSVQSTDAEQFQVNLILR 220
 QY 132 RLKIMNLQOIGRNYNPNPDIDIPSHRLVWFGFTTSILOYENSIMLCTDISHKULRSE 191
 DB 221 RAMEGDLKLVSRYYDPPQAKINLENFRMLQWFGYQTSIRQHENIDILLCEICHKVMRT 280
 QY 192 TVLDFMNFVHOPEEHKFBQVSKELIGLVILFKYNNKTVYVDIDQDNKSTFKADG 251
 DB 281 TLYNLSDAIRDSD--YQSTFRVANGVILTDYNNKTRIDDDVQSTPLCKFTNDG 338
 QY 252 SEVSFLYKRYKQYNOBITDLKQVLVSPQKRRPGGTLFGPAMLIPELCYLTGLTKMKR 311
 DB 339 -EISYVDYKRYNIILRLKQPLVMSRPDKNIRGN-DQAIMIPELARATGWTDAKR 396
 QY 312 NPEVMKDLAVHTLRFPEQRQREVGRLDIYHKNQVQRELWDGLSFDNLLSFGSRL 371

DB 397 ADFTLRAMSEHTRLPDRIRLERLMFNKRLSKCKSQSVETLKSMNIELDSALVEIPARVL 456
 QY 372 QTEKIHQGGKTFDYNPQFADWSKETRGAPLISVPLDNWLLIYTRRNYEAANSLQNLFPK 431
 DB 457 PPEKILFGNQKIFVCDARADWNEFRCTSMFNKHINRWYVITPSRNLRETOEFVQMCIR 516
 QY 432 VTPAMGOMRKAIMIEV--DDTEAYLRVLQOKVTADTQIVVCLLSSNRKDKYDAIKKYL 490
 DB 517 TASSMKMNICPIYBEIIPDRNGTYQAIDNAAANDPQIVMVMVRSPNEEKYSCKIKRKC 576
 QY 491 TDCPTPSQCVVARTILGKQOT----VMAITKIALOMCKMGELWRVDIPLKLVIMVIGD 546
 DB 577 VDRFVPSQVVTLKVIAPRQKPTGLMSIATKVIQMNAKMGAPQVVIPLHGLMTVGFD 636
 QY 547 CVHDMTAGRRSIAGFVASINEGNT--RWFSCIFQDRGOELVDGLKVCILQALRAWNSNE 605
 DB 637 VCHSPKKNKSYGAFVATMDQKESFRYFSTVNEHIKGOELSEQMSVNWACALRSYQEOHR 696
 QY 606 YNPSRIIVRDGVGDGQKLTLYNVEVPQFLDCL----KSIGRGYNPRLTVIVVKRVNTR 661
 DB 697 SLPERILFRDGVGDGQLYQVNVSEVNTLKDRLDIYKSAGKQEGCRMTFIIVSKRINSR 756
 QY 662 PFAQSGGRLQNLPGCTVIDVEVTRPEWYDFIVSOAVRSVSUTHYVIVYDMSGKPDH 721
 DB 757 YF--TGHR--NPVPGTVVDDVITLPERYDFFLVSOAVRIGTVSPTSYVIVISDNMGLNADK 812
 QY 722 IQRITYKLCHYYNMPGVIRVPAPQYAHKLAFLVQSIHREPNLISLNRLYYL 775
 DB 813 LQMSYKTHMYNYSGVIRVPAVCHYAKLAFLVAESINRAPSAGLQNLQYFL 866
 RESULT 5
 ABB62084
 ID ABB62084 standard; Protein; 843 AA.
 XX AC ABB62084;
 XX DT 26-MAR-2002 (first entry)
 XX DE Drosophila melanogaster polypeptide SEQ ID NO 13044.
 XX KW Drosophila; developmental biology; cell signalling; insecticide;
 XX KW pharmaceutical.
 XX OS Drosophila melanogaster.
 XX PN WO200171042-A2.
 XX PD 27-SEP-2001.
 XX PF 23-MAR-2001; 2001WO-US09231.
 XX PR 23-MAR-2000; 2000US-191637P.
 XX PR 11-JUL-2000; 2000US-0614150.
 XX PA (PEKE) PE CORP NY.
 XX PI Venter JC, Adams M, Li PWD, Myers EW;
 XX PX WPI; 2001-656860/75.
 XX DR N-PSDB; ABL06187.
 XX PT New isolated nucleic acid detection reagent for detecting 1000 or more
 XX PT genes from Drosophila and for elucidating cell signalling and cell-cell
 XX PT interactions -
 XX PS Disclosure; SEQ ID NO 13044; 21pp + Sequence Listing; English.
 XX CC The invention relates to an isolated nucleic acid detection reagent
 XX CC capable of detecting 1000 or more genes from Drosophila. The invention is
 XX CC useful in developmental biology and in elucidating cell signalling and
 XX CC cell-cell interactions in higher eukaryotes for the development of


```

317 MDLAVNTRLTLPBORQREVGRLDIYIHKNDVQ---RELRDWGLSFDNSMLLSFSGRILQT 317
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
384 MRAMSSYTRMNPQR---TDLRAFNHRLQNTPSVSKVLRDWNMELDKNVEVQGRITCQ 440
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
374 EKI--HGGKTFDYNPQADWSKETRGAPLLSVKP--LDNMLLIIVTRENYEAAASLIQNL 429
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
441 QNIVFHNGKVPAGEN---ADQWRHFRDQRMILTPSDGLDRWAVIAPQRNSHELRTLLDSL 497
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
430 FKVTPAMGMQMRK-AIMIEVDTRTEAYLRVLQCKVTADTQIVVCILSSNRKDKYDAIKKY 488
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
498 YRAASGMGLRIRSPQEFIIYDRGTGYVRAMDDCVSRDPKLLILCLVPNDNAERYSSIKKR 557
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
489 LCTDPTSPQCVARTLGKQOTVMAIATKIALQWCKMGGELWRVDIPLKLVMIVGIDCY 548
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
558 GYVDRAVPTQVVTLLTKT-KNRSLSMSIATKIALQNLCKLGYPFWMIELPLSLGLMTTGFDA 616
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
549 HDMTAGRSIAGFVASIN-EGMTWFSRCIFQDRGQELVDGLKVCLOAALRAWNSCNEVM 607
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
617 KSTRDRKRAYGALLASMDLQNSYFVSIVTECSAFDVLANTLWPMIAKALROYQHEHRKL 676
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
608 PSRIIVYRDGVDGQLKTLVNYEVPQFDLCKLSIGRGY-----NPRLTIVIVYKKRVNTR 661
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
677 PSRIIVYRDGVSSGSLKQLFEPFVKDIIETKLT---EYARVQLSPQLAYIVVTRSMNTR 733
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
662 FFAQSGGRLQNLPGTVIDVETPEWYDFPIVSAVRSQVSPHYNVIVDNSGLKPDH 721
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
734 FFLNG---QNPPPTGIVDDVITLPERYDFYLVSQQVRQGTVPSTSYNVLYSSMGXSPEK 789
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
722 IQRITYKLCHYYNWPVIRVPAPCOYAKHLAFVIGQSIHREPNLISLRLAYVL 775
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
790 MQKLTYKACHLYYNSGTRVPVAVCOYAKKLATLVGTNLHSLIPQNALEKKFYVL 843
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :

RESULT 7
AAU07866
ID AAU07866 standard; Protein; 580 AA.
XX
AC AAU07866;
XX
DT
DT
DE
DE
XX
XX
KW Mammalian; reproductive-specific protein; male infertility;
KW spermatogenesis; sperm count disorder; anti infertility; reproduction.
XX
OS Mammalia.
XX
FN WO200166752-A2.
XX
PD 13-SEP-2001.
XX
PF 07-MAR-2001; 2001WO-US07371.
XX
XX
PR 07-MAR-2000; 2000US-0187518.
PR 12-JAN-2001; 2001US-0261557.
XX
PA
PA
XX
XX (WHEd ) WHITEHEAD INST BIOMEDICAL RES.
XX
PI Wang PJ, Page DC;
XX
XX WPI; 2001-570774/64.
XX
DR N-PSDE; AAS13630.
DR

```

	reproductive-specific proteins are useful for diagnosing infertility which is a result of reduced sperm count, reduced sperm motility, malformed sperm or combinations of these. The sequences of the invention are useful as markers for spermatogonial cells, for identifying genes or proteins characteristic of male infertility, diagnosing or aiding in the diagnosis of infertility in men, and for contraception in which sperm production or sperm count is reduced or defective sperm is produced. Antibodies to reproductive-specific proteins are useful for determining the presence of these proteins in a sample obtained from a man being assessed for infertility, for identifying the expression of genes in particular cell type or particular developmental stage, for studies of spermatogenesis, and for immunofluorescence of germ cells or in Western blots for assessing the presence of the protein the antibody binds. The sequences of the invention are also useful for treating disorders of reduced sperm count, and for increasing sperm count and/or sperm activity. The nucleic acids of the invention are useful in gene therapy. AAU07859-AAU07882 represent the mammalian reproduction-specific proteins of the present invention.
XX	
SQ	Sequence 580 AA;
	Query Match 33.9%; Score 1389, 5; DB 22; Length 580; Best Local Similarity 44.2%; Pred. No. 3e-126; Matches 258; Conservative 118; Mismatches 199; Indels 9; Gaps 8
QY	197 MFNFYQTBEHKPQEVSKEELIGLVLTKYNKTYRVDDIDWDQNPKSTFKADGSEVSF 256
Db	1 MHAIYQQNEH-FQDECSKLIVGSIVITRYNNRTYRIDVDVNKTPKDSFWMSDGEKITP 59
QY	257 LEYRKQYNQEIIDLKQPVLVSOP-KRRRGPGGTLPGPAMLIPELCYLTLGLTDQRNDFN 315
Db	60 LEYSXNYGITVKEDDQPLLIHRPSERQNNHGMLLKGEILLPELSFWTGPENMKKKDFR 119
QY	316 VMKDIAVHFRLTEPQREVGRLIDIYHKNDNVQRELRDWGLSFDNSLLFSGRILQTEK 375
Db	120 AMKDLTQQINSLSPKHGALCECLLORISQNETASNELTRWGLSHKOVHKLEGRLPMER 179
QY	376 IHOGGKTEDYNPFADWSKE-TSGAPLISKPELDNLWLITERNVEAANSIQLNLFKVP 434
Db	180 INUNTSF-VTSEGILNWKEVTDRDASLTILT-PHFHWALFTPRAMDQARELVNMLEKIAG 237
QY	435 AMGQWRKAIMIEV-DDRTAYLRVLQQ--KWTA DTQIVVCLLSSNRKDKYDAIKKYLC T 491
Db	238 PIGWRTPPAWVELKDDRITYETIRTQSLGVGEKI QMWVCIIINGTRDDLXGAIKKLCV 297
QY	492 DCTPTSCVVARTLGGKOOTVMJAATKIALQNCXMGSELMRVDIPLKLVMITVGIDCYHDM 551
Db	298 QSPVPSQVINVRTIGQPTRLSVAQKILLQWNCCKLGSELWGVDIP LKQLMVIMGDVYHP 357
QY	552 TAGRSIAGHVASINEGMWFNSRCITPDQRGOELVDGLKYCLOAAIRAWNSCNEYMPSEI 611
Db	358 SRGWSRVGVGAVSINITLTKWYGRVFPQMPOHEITVDSLKGLCVGLSKKYEYEVNHCLPEKI 417
QY	612 IVYRDGVGDQLKATLVNVEYPQFLDCLKSTGRGNRPRLTVIVKKRVNMTREFFAOSGGR LQ 671
Db	418 VVTRDGVSQDLKVANYEIPOLQKFEAFD-NYHPKMVFVVOKKISTNLYLAAPDFV 476
QY	672 NPLPGTVIDVEVTRPEWYDFPIVQSQA VRSGSVSPTHFYINVYDNSGLKPDIHTQLRTYKLC H 731
Db	477 TPSEGTIVDHTITSCEWVD FYLLAHVRVCGGPIPTHYICVINTANISP DHMQRTFLKCH 536
QY	732 IYYNWGVTVRPAPCOYAHKLAFLVGOSSHREPNLSLNRIYYL 775
Db	537 MYWNWGTTRVPAPCYAHKLAFLSGQILHHEPAIQLCGNLFFL 580

xx DE Human mRNA sequence.

xx KW Neuroprotective; immunomodulator; cancer;

xx KW cytoskeletal; anti-inflammatory; gene therapy; nutritional supplement;

xx KW wound; burn; Alzheimer's disease; Huntington's disease;

xx KW amyotrophic lateral sclerosis; autoimmune disorder; inflammation;

xx KW vulnary.

xx OS Homo sapiens.

xx PN WO200231111-A2.

xx PD 18-APR-2002.

xx PF 11-OCT-2001; 2001WO-US27760.

xx PR 12-OCT-2000; 2000US-0687527.

xx PA (HYSE-) HYSEQ INC.

xx PI Tang YT, Liu C, Zhou P, Asundi V, Zhang J, Zhao QA, Ren F;

xx PI Xue AJ, Yang Y, Wehrman T, Drmanac RT;

xx DR WPI; 2002-426278/45.

xx DR N-PSDB; AB061109.

xx XX New polypeptides and their encoded proteins, useful as nutritional

PT sources or supplements, or in gene therapy, particularly for treating

PT wounds, Alzheimer's disease, amyotrophic lateral sclerosis, cancer or

PT inflammation -

xx PS Claim 20; SEQ ID # 768; 357pp + sequence listing; English.

xx CC The invention relates to 446 newly isolated polynucleotide sequences.

CC The activity of polynucleotides of the invention may be described as,

CC vulnary, neuroprotective, immunomodulator, cytoskeletal and

CC anti-inflammatory. Compositions comprising nucleic acids of the invention

CC are useful for treating a mammalian subject, or as nutritional sources or

CC supplements. These are useful in gene therapy, particularly for treating

CC wounds, burns or ulcers, Alzheimer's disease, Huntington's disease,

CC amyotrophic lateral sclerosis, autoimmune disorders, cancer or

CC inflammation. The nucleic acids and polypeptides are also useful in

CC diagnostic and research methods. The sequences, given in records

CC ABP43544-ABP43989 represent polypeptides encoded by polynucleotides of

CC the invention.

CC NOTE: The sequence data for this patent did not form part of the printed

CC specification, but was obtained in electronic format directly from WIPO

CC at ftp.wipo.int/pub/published_pct_sequences.

xx SQ Sequence 498 AA;

Query Match 32.3%; Score 1324.5; DB 23; Length 498;

Best Local Similarity 51.4%; Pred. No. 5.3e-120;

Matches 259; Conservative 86; Mismatches 150; Indels 9; Gaps 4;

QY 275 VLVSQKRGCGTLPAGMILPELCYLTGTDMKNDNFNMKDLAVHTRLTPEQORE 334

DB 1 MLVSLKKKRNDSN-PQLAHLPELCFTLTQATSDPQLMKAVAEKTRLSFGROQR 59

QY 335 VGLIDYIHKNDNVQRELDWGLSFDNLSLFSGRILQTEKHGGKTFDYNPQ---FAD 391

DB 60 LARLVNIOQNTWAPETWGLHFGSQ-ISLGRIVPSEKILMQ---DHICQPVSAAD 114

QY 392 WSKETRGAFLISVKPLDNMLLIYTRRNYEAANSLIQNLFKVTPAGMOMKRAIMIEVDDR 451

DB 115 WSKDIKTLNQAQSLNTWLILCSDRTEYVAESFNLCLRRVAGSMGFNDYPKIIKQVEN 174

QY 452 TEAYLRVLOQKVTADTQIVVCLLSNRKOKYDAIKKYLCTDCPTSCQCVVARTLHGQQT 511

DB 175 PRAFVRAIQVYDPDVLQVLMCILFSPKQTYDSIKKYLSSDCVPSPQCVLARTLNQGMW 234

QY 512 MATATKIALQNMCKMGELWRVDIPLKLVMIIVGIDCYHDMTAGRRSIAGFVASINSGMTR 571

Db 235 MSATKIAMQMTCKLGGELWAVEIPKLSLMVVGIDVCKDALSCKDVMVGCVASVNPTRITR 294

QY 572 WFSRCIFQDRGOELVDGLKVCLOAALRAWNSCNEYMPRIIVYRDGCGGOLKTLVNYEV 631

Db 295 WFSRCILQRTWTDVADCLKVFMTGALNKWYKYNHDLPARIIIVYRAGDGGOLKTLIEYV 354

QY 632 PQFLDCLKSIGRGYNPRLTVIWKRVNTRFFAOSGGRLOQNPFGTGVIVDVEVTRPEWYDF 691

Db 355 PQLSSVAESSNTSSRLSVIVRKKCMRPFTEHNRVQNPPLGTGVVDSSEATRNWYDF 414

QY 692 FIVSQAVRSGSVSPHYNVYDNSGLKPDHITQRTYKLCIHYNNWPGVIRVPAPQOYAHK 751

Db 415 YLISQVACRGTVSPTYNNVIYDNGLKFEDHMQRLTFKLCHLYNNWPGVIRVPAPQOYAHK 474

QY 752 LAFVQGSIHREPNSLSNRLYYL 775

Db 475 LTFIGAQSIIHKEPSLELANHLFYL 498

RESULT 9

AAB94209

ID AAB94209 standard; Protein; 530 AA.

XX AAB94209;

AC AAB94209;

XX 26-JUN-2001 (first entry)

XX Human protein sequence SEQ ID NO:14559.

DE Human; primer; detection; diagnosis; antisense therapy; gene therapy.

XX Homo sapiens.

XX EP1074617-A2.

XX 07-FEB-2001.

XX 28-JUL-2000; 2000EP-0116126.

XX 29-JUL-1999; 99JP-0248036.

XX 27-AUG-1999; 99JP-0300253.

XX 11-JAN-2000; 2000JP-0118776.

XX 02-MAY-2000; 2000JP-0183767.

XX 09-JUN-2000; 2000JP-0241899.

XX (HELI-) HELIX RES INST.

XX Ota T, Isogai T, Nishikawa T, Hayashi K, Saito K, Yamamoto J;

PI Ishii S, Sugiyama T, Wakamatsu A, Nagai K, Otsuki T;

XX WPI; 2001-318749/34.

XX Primer sets for synthesizing polynucleotides, particularly the 5602

PT full-length cDNAs defined in the specification, and for the detection

PT and/or diagnosis of the abnormality of the proteins encoded by the

PT full-length cDNAs -

XX Claim 8; SEQ ID 14559; 2537pp + CD ROM; English.

XX The present invention describes primer sets for synthesizing 5602

CC full-length cDNAs defined in the specification. Where a primer set

CC comprises: (a) an oligo-dT primer and an oligonucleotide complementary

CC to the complementary strand of a polynucleotide which comprises one of

CC the 5602 nucleotide sequences defined in the specification, where the

CC oligonucleotide comprises at least 15 nucleotides; or (b) a combination

CC of an oligonucleotide comprising a sequence complementary to the

CC complementary strand of a polynucleotide which comprises a 5'-end

CC sequence and an oligonucleotide comprising a sequence complementary to a

CC polynucleotide which comprises a 3'-end sequence, where the

CC oligonucleotide comprises at least 15 nucleotides and the combination of

CC the 5'-end sequence/3'-end sequence is selected from those defined in

CC the specification. The primer sets can be used in antisense therapy and

QY 767 SLSNRLYL 775
 Db 363 QLCENLFEL 371
 RESULT 11
 AAB58977
 ID AAB58977 standard; Protein; 130 AA.
 XX
 AC AAB58977;
 XX
 DT 27-MAR-2001 (first entry)
 XX
 DE Breast and ovarian cancer associated antigen protein sequence SEQ ID 685.
 XX
 KW Human; breast cancer; ovarian cancer; cytostatic; immunosuppressive;
 KW nontropic; neuroprotective; antiviral; antiallergic; hepatotropic;
 KW antidiabetic; antiinflammatory; anticancer; vulnar; anticonvulsant;
 KW antibacterial; antifungal; antiparasitic; cardiant; immune disorder;
 KW Addison's disease; allergy; autoimmune haemolytic anaemia;
 KW autoimmune thyroiditis; diabetes mellitus; Crohn's disease;
 KW multiple sclerosis; rheumatoid arthritis; ulcerative colitis;
 KW cardiovascular disorder; wound healing; neurological disease.
 XX
 OS Homo sapiens.
 XX
 PN WO200055173-A1.
 XX
 PD 21-SEP-2000.
 XX
 PF 08-MAR-2000; 2000WO-US05881.
 XX
 PR 12-MAR-1999; 99US-0124270.
 XX
 PA (HUMA-) HUMAN GENOME SCI INC.
 XX
 PI Rosen CA, Ruben SM;
 XX
 DR WPI; 2000-611515/58.
 DR N-PSDB; AAF21880.
 XX
 PT New human breast and ovarian cancer associated gene sequences and the
 PT polypeptides encoded by these genes, useful in the prevention,
 PT treatment and diagnosis of cancer, immune disorders, cardiovascular
 PT disorders and neurological diseases -
 PS
 PS Claim 11; Page 1139; 1299pp; English.
 CC
 CC Sequences AAF21614 - AAF22031 represent DNA sequences encoding human
 CC proteins AAB58711 - AAB59128. The DNA and protein sequences are
 CC associated with breast and ovarian cancer. Included in the invention are
 CC sequences AAF22032 - AAF22040 and AAB59129 which are used in the
 CC isolation and characterisation of the DNA and protein sequences of the
 CC invention. The breast and ovarian cancer associated DNA, protein, agonist
 CC or antagonist sequences exhibit cytostatic; immunosuppressive;
 CC nontropic; neuroprotective; antiviral; antiallergic; hepatotropic;
 CC antidiabetic; antiinflammatory; anticancer; vulnar; anticonvulsant;
 CC antibacterial; antifungal; antiparasitic and cardiant activity. The
 CC polynucleotide and protein sequences are used in the diagnosis of cancer,
 CC particularly breast and ovarian cancer. The nucleic acid sequences,
 CC proteins, agonists and antagonists may also be used in the diagnosis,
 CC prevention and treatment of immune disorders e.g. Addison's disease,
 CC allergies, autoimmune haemolytic anaemia, autoimmune thyroiditis,
 CC diabetes mellitus, Crohn's disease, multiple sclerosis, rheumatoid
 CC arthritis and ulcerative colitis; cardiovascular disorders such as
 CC myocardial ischaemias; wound healing; neurological diseases such as
 CC cerebral anoxia and epilepsy; and infectious diseases.
 CC
 SQ Sequence 130 AA;
 Query Match 15.9%; Score 652; DB 21; Length 130;
 Best Local Similarity 99.2%; Pred. No. 3.2e-55;
 Matches 123; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 492 DCPTSCQCVARTLGKQOTVMAIKIALQWNCXMGGLWVRVDIELKLWMIYIGIDCYHDM 551
 Db 5 DCPTSCQCVARTLGKQOTVMAIKIALQWNCXMGGLWVRVDIELKLWMIYIGIDCYHDM 64
 QY 552 TAGRSIAGFVASINEGMRWFSRCIFODRGQELVDGLKVCLOAALRAWNSCNEYMPRSI 611
 Db 65 TAGRSIAGFVASINEGMRWFSRCIFODRGQELVDGLKVCLOAALRAWNSCNEYMPRSI 124
 QY 612 IVYR 615
 Db 125 IVYR 128
 RESULT 12
 AAG05957
 ID AAG05957 standard; Protein; 341 AA.
 XX
 AC AAG05957;
 XX
 DT 13-FEB-2002 (first entry)
 XX
 DE Novel human diagnostic protein #5948.
 XX
 KW Human; chromosome mapping; gene mapping; gene therapy; forensic;
 KW food supplement; medical imaging; diagnostic; genetic disorder.
 XX
 OS Homo sapiens.
 XX
 PN WO200175067-A2.
 XX
 PD 11-OCT-2001.
 XX
 PF 30-MAR-2001; 2001WO-US08631.
 XX
 PR 31-MAR-2000; 2000US-0540217.
 PR 23-AUG-2000; 2000US-0649167.
 XX
 PA (HYSE-) HYSEQ INC.
 XX
 PI Drmanac RT, Liu C, Tang YT;
 XX
 DR WPI; 2001-639362/73.
 DR N-PSDB; AAS70144.
 XX
 PT New isolated polynucleotide and encoded polypeptides, useful in
 PT diagnostics, forensics, gene mapping, identification of mutations
 PT responsible for genetic disorders or other traits and to assess
 PT biodiversity -
 PT
 PT Claim 20; SEQ ID No 36316; 103pp; English.
 CC
 CC The invention relates to isolated polynucleotide (I) and
 CC polypeptide (II) sequences. (I) is useful as hybridisation probes,
 CC polymerase chain reaction (PCR) primers, oligomers, and for chromosome
 CC and gene mapping, and in recombinant production of (II). The
 CC polynucleotides are also used in diagnostics as expressed sequence tags
 CC for identifying expressed genes. (I) is useful in gene therapy techniques
 CC to restore normal activity of (II) or to treat disease states involving
 CC (II). (II) is useful for generating antibodies against it, detecting or
 CC quantitating a polypeptide in tissue, as molecular weight markers and as
 CC a food supplement. (II) and its binding partners are useful in medical
 CC imaging of sites expressing (II). (I) and (II) are useful for treating
 CC disorders involving aberrant protein expression or biological activity.
 CC The polypeptide and polynucleotide sequences have applications in
 CC diagnostics, forensics, gene mapping, identification of mutations
 CC responsible for genetic disorders or other traits to assess biodiversity
 CC and to produce other types of data and products dependent on DNA and
 CC amino acid sequences. AAG00010-AAG30377 represent novel human
 CC diagnostic amino acid sequences of the invention.
 CC Note: The sequence data for this patent did not appear in the printed
 CC specification, but was obtained in electronic format directly from WIPO
 CC at ftp.wipo.int/pub/published_pct_sequences.

XX SQ Sequence 341 AA;
Query Match 14.1%; Score 580; DB 22; Length 341;
Best Local Similarity 43.5%; Pred. No. 1.8e-47;
Matches 114; Conservative 50; Mismatches 64; Indels 34; Gaps 3;
QY 89 QQKTEVFKTRNGEDVRITITLNLPTSPPTCLOFYNIIFRLILKIMNLOQIGENYIN 148
Db 16 QPTVTELSSETORGETIKTITLKRELPSPPVICQVENIIFR-----58
QY 149 PNDPIDIPHLVINGFTTSILQYENSIMLCITDVSHKVLRSYTDPMFNPHQTEEHK 208
Db 59 -----KLSLWPGFAISVYFRKLIFSADVSQVLRNETVLEFMTALCORTGLSC 108
QY 209 FQSOVSKELIGLVLYKNNKTYRVDIDWDQPKSTFEKADGSEVSELEYRKOYNQEI 268
Db 109 FTCTCKQLLGLVLYRNNRTYSIDIDMSVPTTFQKRDGTETIYVDIYKQDYDITV 168
QY 269 TDLKQPVLSQPKRRRPGGTLFPGPAMLIPELCYLTGLTDKMDNFVNMKDLAVHTRLTP 328
Db 169 SDLNQPMVLSLLKKKNDNSEAQ-LAHLIPELCFLGLTDQATSDFLMKVAEKTRLSP 227
QY 329 EQRQREVGLIDVIHKNDNVQR 350
Db 228 SCRQQRRLRV-----DNRIQR 243
RESULT 13
AAG42119
ID AAG42119 standard; Protein; 869 AA.
XX AC AAG42119;
XX AC AC
XX DT 18-OCT-2000 (first entry)
XX DE Arabidopsis thaliana protein fragment SEQ ID NO: 52488.
XX KW Protein identification; signal transduction pathway; metabolic pathway;
XX KW hybridisation assay; genetic mapping; gene expression control; promoter;
XX KW termination sequence.
XX OS Arabidopsis thaliana.
XX PN EP1033405-A2.
XX PD 06-SEP-2000.
XX PF 25-FEB-2000; 2000EP-0301439.
XX PR 25-FEB-1999; 99US-0121825.
XX PR 05-MAR-1999; 99US-0123180.
XX PR 09-MAR-1999; 99US-0123548.
XX PR 23-MAR-1999; 99US-0125788.
XX PR 25-MAR-1999; 99US-0126264.
XX PR 29-MAR-1999; 99US-0126785.
XX PR 01-APR-1999; 99US-0127462.
XX PR 06-APR-1999; 99US-0128234.
XX PR 08-APR-1999; 99US-0128714.
XX PR 16-APR-1999; 99US-0129845.
XX PR 19-APR-1999; 99US-0130077.
XX PR 21-APR-1999; 99US-0130449.
XX PR 23-APR-1999; 99US-0130510.
XX PR 28-APR-1999; 99US-0130891.
XX PR 30-APR-1999; 99US-0131449.
XX PR 30-APR-1999; 99US-0132048.
XX PR 04-MAY-1999; 99US-0132484.
XX PR 05-MAY-1999; 99US-0132485.
XX PR 06-MAY-1999; 99US-0132486.
XX PR 06-MAY-1999; 99US-0132487.
XX PR 07-MAY-1999; 99US-0132863.
XX PR 11-MAY-1999; 99US-0134256.
PR 14-MAY-1999; 99US-0134218.
PR 14-MAY-1999; 99US-0134219.
PR 14-MAY-1999; 99US-0134221.
PR 14-MAY-1999; 99US-0134370.
PR 18-MAY-1999; 99US-0134768.
PR 19-MAY-1999; 99US-0134941.
PR 20-MAY-1999; 99US-0135124.
PR 21-MAY-1999; 99US-0135353.
PR 24-MAY-1999; 99US-0135629.
PR 25-MAY-1999; 99US-0136021.
PR 27-MAY-1999; 99US-0136392.
PR 28-MAY-1999; 99US-0136782.
PR 01-JUN-1999; 99US-0137222.
PR 03-JUN-1999; 99US-0137528.
PR 04-JUN-1999; 99US-0137502.
PR 07-JUN-1999; 99US-0137724.
PR 08-JUN-1999; 99US-0138094.
PR 10-JUN-1999; 99US-0138540.
PR 10-JUN-1999; 99US-0138847.
PR 14-JUN-1999; 99US-0139119.
PR 16-JUN-1999; 99US-0139452.
PR 16-JUN-1999; 99US-0139453.
PR 17-JUN-1999; 99US-0139492.
PR 18-JUN-1999; 99US-0139454.
PR 18-JUN-1999; 99US-0139455.
PR 18-JUN-1999; 99US-0139456.
PR 18-JUN-1999; 99US-0139457.
PR 18-JUN-1999; 99US-0139458.
PR 18-JUN-1999; 99US-0139459.
PR 18-JUN-1999; 99US-0139460.
PR 18-JUN-1999; 99US-0139461.
PR 18-JUN-1999; 99US-0139462.
PR 18-JUN-1999; 99US-0139463.
PR 18-JUN-1999; 99US-0139750.
PR 18-JUN-1999; 99US-0139763.
PR 21-JUN-1999; 99US-0139817.
PR 22-JUN-1999; 99US-0139899.
PR 23-JUN-1999; 99US-0140353.
PR 23-JUN-1999; 99US-0140354.
PR 24-JUN-1999; 99US-0140695.
PR 28-JUN-1999; 99US-0140823.
PR 29-JUN-1999; 99US-0140991.
PR 30-JUN-1999; 99US-0141287.
PR 01-JUL-1999; 99US-0141842.
PR 01-JUL-1999; 99US-0142154.
PR 02-JUL-1999; 99US-0142055.
PR 06-JUL-1999; 99US-0142390.
PR 08-JUL-1999; 99US-0142803.
PR 09-JUL-1999; 99US-0142920.
PR 12-JUL-1999; 99US-0142977.
PR 13-JUL-1999; 99US-0143542.
PR 14-JUL-1999; 99US-0143624.
PR 15-JUL-1999; 99US-0144005.
PR 16-JUL-1999; 99US-0144085.
PR 16-JUL-1999; 99US-0144086.
PR 19-JUL-1999; 99US-0144325.
PR 19-JUL-1999; 99US-0144331.
PR 19-JUL-1999; 99US-0144332.
PR 19-JUL-1999; 99US-0144333.
PR 19-JUL-1999; 99US-0144334.
PR 19-JUL-1999; 99US-0144335.
PR 20-JUL-1999; 99US-0144352.
PR 20-JUL-1999; 99US-0144632.
PR 20-JUL-1999; 99US-0144884.
PR 21-JUL-1999; 99US-0144814.
PR 21-JUL-1999; 99US-0145086.
PR 21-JUL-1999; 99US-0145088.
PR 22-JUL-1999; 99US-0145085.
PR 22-JUL-1999; 99US-0145087.
PR 22-JUL-1999; 99US-0145089.
PR 22-JUL-1999; 99US-0145192.
PR 23-JUL-1999; 99US-0145145.
PR 23-JUL-1999; 99US-0145218.

Db 818 PETSDBGSM 826

RESULT 14
AAG42118
ID AAG42118 standard; Protein; 1048 AA.
XX AC AAG42118;
XX DT 18-OCT-2000 (first entry)
XX DE Arabidopsis thaliana protein fragment SEQ ID NO: 52487.
XX KW Protein identification; signal transduction pathway; metabolic pathway;
KW hybridisation assay; genetic mapping; gene expression control; promoter;
KW termination sequence.
XX OS Arabidopsis thaliana.
XX PN BP1033405-A2.
XX PD 06-SEP-2000.
XX PF 25-FEB-2000; 2000EP-0301439.
XX PR 25-FEB-1999; 99US-0121825.
PR 05-MAR-1999; 99US-0123180.
PR 09-MAR-1999; 99US-0123548.
PR 23-MAR-1999; 99US-0125788.
PR 25-MAR-1999; 99US-0126264.
PR 29-MAR-1999; 99US-0126785.
PR 01-APR-1999; 99US-0127462.
PR 06-APR-1999; 99US-0128234.
PR 08-APR-1999; 99US-0128714.
PR 16-APR-1999; 99US-0129845.
PR 19-APR-1999; 99US-0130077.
PR 21-APR-1999; 99US-0130449.
PR 23-APR-1999; 99US-0130510.
PR 28-APR-1999; 99US-0130891.
PR 30-APR-1999; 99US-0131449.
PR 04-MAY-1999; 99US-0132484.
PR 05-MAY-1999; 99US-0132485.
PR 06-MAY-1999; 99US-0132486.
PR 07-MAY-1999; 99US-0132863.
PR 11-MAY-1999; 99US-0134256.
PR 14-MAY-1999; 99US-0134218.
PR 14-MAY-1999; 99US-0134219.
PR 14-MAY-1999; 99US-0134321.
PR 14-MAY-1999; 99US-0134370.
PR 18-MAY-1999; 99US-0134768.
PR 19-MAY-1999; 99US-0134941.
PR 20-MAY-1999; 99US-0135124.
PR 21-MAY-1999; 99US-0135353.
PR 24-MAY-1999; 99US-0135629.
PR 25-MAY-1999; 99US-0136021.
PR 27-MAY-1999; 99US-0136382.
PR 28-MAY-1999; 99US-0136782.
PR 01-JUN-1999; 99US-0137222.
PR 03-JUN-1999; 99US-0137528.
PR 04-JUN-1999; 99US-0137502.
PR 07-JUN-1999; 99US-0137724.
PR 08-JUN-1999; 99US-0138094.
PR 10-JUN-1999; 99US-0138540.
PR 10-JUN-1999; 99US-0138847.
PR 14-JUN-1999; 99US-0139119.
PR 16-JUN-1999; 99US-0139452.
PR 16-JUN-1999; 99US-0139453.
PR 17-JUN-1999; 99US-0139942.
PR 18-JUN-1999; 99US-0139454.
PR 18-JUN-1999; 99US-0139455.
PR 18-JUN-1999; 99US-0139456.
PR 18-JUN-1999; 99US-0139457.
PR 18-JUN-1999; 99US-0139458.
PR 18-JUN-1999; 99US-0139459.
PR 18-JUN-1999; 99US-0139460.
PR 18-JUN-1999; 99US-0139461.
PR 18-JUN-1999; 99US-0139462.
PR 18-JUN-1999; 99US-0139463.
PR 18-JUN-1999; 99US-0139750.
PR 18-JUN-1999; 99US-0139763.
PR 21-JUN-1999; 99US-0139817.
PR 22-JUN-1999; 99US-0139899.
PR 23-JUN-1999; 99US-0140353.
PR 23-JUN-1999; 99US-0140354.
PR 24-JUN-1999; 99US-0140695.
PR 28-JUN-1999; 99US-0140823.
PR 29-JUN-1999; 99US-0140991.
PR 30-JUN-1999; 99US-0141287.
PR 01-JUL-1999; 99US-0141842.
PR 02-JUL-1999; 99US-0142055.
PR 06-JUL-1999; 99US-0142390.
PR 08-JUL-1999; 99US-0142803.
PR 09-JUL-1999; 99US-0142920.
PR 12-JUL-1999; 99US-0142977.
PR 13-JUL-1999; 99US-0143542.
PR 14-JUL-1999; 99US-0143624.
PR 15-JUL-1999; 99US-0144005.
PR 16-JUL-1999; 99US-0144085.
PR 16-JUL-1999; 99US-0144086.
PR 19-JUL-1999; 99US-0144325.
PR 19-JUL-1999; 99US-0144331.
PR 19-JUL-1999; 99US-0144332.
PR 19-JUL-1999; 99US-0144333.
PR 19-JUL-1999; 99US-0144334.
PR 19-JUL-1999; 99US-0144335.
PR 20-JUL-1999; 99US-0144352.
PR 20-JUL-1999; 99US-0144632.
PR 20-JUL-1999; 99US-0144884.
PR 21-JUL-1999; 99US-0144814.
PR 21-JUL-1999; 99US-0145086.
PR 22-JUL-1999; 99US-0145087.
PR 22-JUL-1999; 99US-0145089.
PR 22-JUL-1999; 99US-0145192.
PR 23-JUL-1999; 99US-0145145.
PR 23-JUL-1999; 99US-0145218.
PR 23-JUL-1999; 99US-0145224.
PR 26-JUL-1999; 99US-0145276.
PR 27-JUL-1999; 99US-0145913.
PR 27-JUL-1999; 99US-0145918.
PR 27-JUL-1999; 99US-0145919.
PR 28-JUL-1999; 99US-0145951.
PR 02-AUG-1999; 99US-0146386.
PR 02-AUG-1999; 99US-0146388.
PR 03-AUG-1999; 99US-0146389.
PR 04-AUG-1999; 99US-0147038.
PR 04-AUG-1999; 99US-0147204.
PR 05-AUG-1999; 99US-0147302.
PR 05-AUG-1999; 99US-0147192.
PR 06-AUG-1999; 99US-0147260.
PR 06-AUG-1999; 99US-0147303.
PR 06-AUG-1999; 99US-0147416.
PR 09-AUG-1999; 99US-0147493.
PR 09-AUG-1999; 99US-0147935.
PR 10-AUG-1999; 99US-0148171.
PR 11-AUG-1999; 99US-0148319.
PR 12-AUG-1999; 99US-0148341.
PR 13-AUG-1999; 99US-0148565.
PR 13-AUG-1999; 99US-0148684.
PR 16-AUG-1999; 99US-0149368.

PR	17-AUG-1999;	99US-0149175.	QY	110	TLTNELPPTGTCLOFYNIIFRRLLKIMNLOOIGRNYNPNNDPIDPSHR-----LVWLP	164
PR	18-AUG-1999;	99US-0149426.	Db	305	FLECKQSDAPQALQVLDIVLRE-LPTSRYIPVGRSFYSP-----DIGKKQLSGDLESWR	359
PR	20-AUG-1999;	99US-0149722.	QY	165	GFTTSIIQYENSIMLCVDVSHKV-LRSETVLDNFNFYHQTEEHK-----PQEQVSKELI	218
PR	20-AUG-1999;	99US-0149929.	Db	360	GFYQSIPTQMLSLNIDMSSTAFIEANPVIFQVCDLLNRDISRPLSDADRVIKKALR	419
PR	23-AUG-1999;	99US-0149902.	QY	219	GL-VVLAKYNN--KTYRVDDIDWDQNPSTPKKAD-GSEVSFLEYRYKOYNOEITDLKOP	274
PR	23-AUG-1999;	99US-0150566.	Db	420	GKVEVTHRGNNRRKYRISGLTAVATRELTPVDERNTQKSVYFHYETFRICHTOLP	479
PR	26-AUG-1999;	99US-0150884.	QY	275	VLSQPKRRRGFGGTLPGFAMLIPELCYLT--GLTDMKNDNFNMKDLAVHRLPPEORQ	332
PR	27-AUG-1999;	99US-0151065.	Db	480	CLQVGSNR-----PNYLPMEVCKIVGQYRSKELNERQITALLKV--TCQRPIDRE	529
PR	27-AUG-1999;	99US-0151303.	QY	333	REV---GRLLIYIHKNDVQBELDWGLSFPSNLLSPSGRILOTE--KIHOGGKTFDYNP	387
PR	30-AUG-1999;	99US-0151303.	Db	530	KDILQTVQLNDY--AKDNYAQE--FGIKISTLASVEARILPPPLWKYHESGREGTCLP	584
PR	31-AUG-1999;	99US-0151438.	QY	388	QFADWSKETRGAPLISVKPLDNLLIYTRNVE--AANSLIQNLFPKVTAPMGQMR--KAI	444
PR	01-SEP-1999;	99US-0151930.	Db	585	QVGQNNMMNK--KMINGTVMNWCINFSRQVQDNLARTFCOELAQMCYVSGMAFNPPV	642
PR	07-SEP-1999;	99US-0152363.	QY	445	MLEVDDRTEAYLRVLQKVTADT-----QIVVCLLSSNRKDKYDAIKYLCTDCTP	495
PR	10-SEP-1999;	99US-0153070.	Db	643	LPPVSARPEQVEKVLKTRYHDATSKLSQGEIDLLIIVLPDNNGLYGLDKRICETELGI	702
PR	13-SEP-1999;	99US-0153758.	QY	496	PSQCVVARTLQK-QQTVMAIATKIALQNCXMGG-ELWRVD-----IPL---KLVNIVGI	545
PR	15-SEP-1999;	99US-0154018.	Db	703	VSQCCLTRKHVFMKQYMA--NVALKINVKVGGNTVLVDALSRRIFLVSDRPTIIFGA	759
PR	16-SEP-1999;	99US-0154039.	QY	546	DCYHDMTA--GRRSTAGFVASIN-BGTRWTFSCIFQDRGOELVDGL-----	589
PR	20-SEP-1999;	99US-0154779.	Db	760	DVTHEPHGEDSPSIAAVVASQDNWPEITKIAGLVCQAHRQELIQLDKPEWDPQKGVVT	819
PR	22-SEP-1999;	99US-0155139.	QY	590	KVCLQALRAMNMCNEYMPRIIVYRDGVGGQLKTLVNYEVPQFLDCLKLSIGRGYNPL	649
PR	24-SEP-1999;	99US-0155659.	Db	820	GGMIKELLIAPFRSTGHKPLRIIFYRDGVSEQFQVQLLYELDLDAIRKACASLEAGYQPPV	879
PR	28-SEP-1999;	99US-0156458.	QY	650	TVIVVKRVNTRFFAQ-----SGRLQNLPGTVIDVEVPEWYDFPIVQAVRSQSV	703
PR	29-SEP-1999;	99US-0156596.	Db	880	TFVVYQKREHTLFAQENDRHSVDRSGNILPTGVDSKICHTPTEFDFYLCSHAGIQGTS	939
PR	04-OCT-1999;	99US-0157117.	QY	704	SPTHYNIYDNGSLGKPDHITQRTLYKLIYNNWPGVIRVAPCQYAHKLAFLVGQSIHRE	763
PR	05-OCT-1999;	99US-0157753.	Db	940	RPAHYHVLWDENNFTADGLQSLTNLCTIYARCTRSVSIVPPAYAHLAFAFA--RPFYME	997
PR	06-OCT-1999;	99US-0157865.	QY	764	PNLSLSNRL 772	
PR	07-OCT-1999;	99US-0158029.	Db	998	PETSDSGSM 1006	
PR	08-OCT-1999;	99US-0158232.	QY	RESULT 15		
PR	12-OCT-1999;	99US-0158369.	Db	ABP64718		
PR	13-OCT-1999;	99US-0159293.	QY	ID	ABP64718 standard; Protein; 859 AA.	
PR	13-OCT-1999;	99US-0159294.	Db	XX	ABP64718;	
PR	14-OCT-1999;	99US-0159329.	QY	AC	ABP64718;	
PR	14-OCT-1999;	99US-0159330.	Db	XX	25-FEB-2003 (first entry)	
PR	14-OCT-1999;	99US-0159331.	QY	Human protein SEQ ID 378.		
PR	14-OCT-1999;	99US-0159637.	Db	Human; expressed sequence tag; EST;		
PR	14-OCT-1999;	99US-0159638.	QY	KW	haematopoietic disorder; central nervous system disease; viral infection;	
PR	18-OCT-1999;	99US-0159584.	Db	KW	peripheral nervous system disease; non-healing wound; infectious disease;	
PR	21-OCT-1999;	99US-0160741.	QY	KW	immune deficiency; immune disorder; bacterial infection; allergy; cancer;	
PR	21-OCT-1999;	99US-0160767.	Db	KW	fungal infection; autoimmune disorder; coagulation disorder; neutropenic;	
PR	21-OCT-1999;	99US-0160768.	QY	KW	antiallergic; antiinflammatory; immunosuppressive; neuroprotective;	
PR	21-OCT-1999;	99US-0160770.	Db	KW	cytostatic; haemostatic; virucide; antibacterial; fungicide;	
PR	21-OCT-1999;	99US-0160814.	QY	XX	immunostimulant; cerebroprotective.	
PR	21-OCT-1999;	99US-0160815.	Db	OS	Homo sapiens.	
PR	22-OCT-1999;	99US-0160980.	QY			
PR	22-OCT-1999;	99US-0160981.	Db			
PR	22-OCT-1999;	99US-0160989.	QY			
PR	22-OCT-1999;	99US-0161404.	Db			
PR	25-OCT-1999;	99US-0161405.	QY			
PR	25-OCT-1999;	99US-0161406.	Db			
PR	26-OCT-1999;	99US-0161359.	QY			
PR	26-OCT-1999;	99US-0161360.	Db			
PR	26-OCT-1999;	99US-0161361.	QY			
PR	28-OCT-1999;	99US-0161320.	Db			
PR	28-OCT-1999;	99US-0161992.	QY			
PR	28-OCT-1999;	99US-0161993.	Db			
PR	29-OCT-1999;	99US-0162142.	QY			

Query Match 12.4%; Score 510; DB 21; Length 1048;

Best Local Similarity 25.1%; Pred. No. 8.2e-40; Matches 213; Conservative 136; Mismatches 376; Indels 124; Gaps 33;

QY	20	GSSGIIIVRLSTNFRSLTRPQWALYQHYDYNPLMEARRLSALLFOHEDLICKCH----	75
Db	186	GQSGKRCIVKANHP-FALPDKDLHYDVITTEVTSRGVNRVAMKQLVDNRYDSHLGSR	244
QY	76	--AFDG-----TILFLP-----KRLQKQTEVFSKTRNGEDVRITI	109
Db	245	LPAVDGRKSLYTAGPLFPFNSKEPRINLLDEVGAGGQRREREFKVIKVARADLHLGLM	304

[illegible]

269	Db	ITHCGMKRYRVCNVTTRFPASHQTPELQOESGCTVECTVAQYFKDRHKLVLRYPHPLCL	328
277	QY	VSPQKRRRGGTLPGLMILPELCLVTG-----LTDKMNENDNMKDLAVHTRLP	328
329	Db	QVGEQKH-----TVLPLEVCNIVAGORCIKKLTD--NQSTM--IRATARSAP	373
329	QY	FORQEVGRGLIDYTHKNDNVORELRDGLSFDNLLSPSGRILQTEKKIHQGGKTFDYNFQ	388
374	Db	D-ROEISIKLMR--SASFNTDPYVREFGIMVKDEMDVTGRVLQPPSIILYQGR-----	423
389	QY	FADWSKETRGAPLISVKPELN-----MLIIVTRNEYAA---NSLIQNLFKVT	433
424	Db	-----NKAIATPVGVGWMNRNKPHTGIBIKVAIACFAPQROCTEVHHLKSFTEQURKIS	478
434	QY	PAMGQMEK-----AIMIEVDRTDEAYLRVLQOQKVTADTQIVVCLLSNRKDYDAIKKYL	489
479	Db	RDAGMPIQGOFCPKYAGGADSVZPEMERHL-KNTYAGLQVLVVIL-PGKTPVYAEVKRGV	536
490	QY	CTDCTPQOCVVARTLKGQOTWAIATKIALQMNCKMG-----ELWRVDTPLKLVIMV	543
537	Db	DTVLGMATQCQMKNV--QRTPTQSLNCLGKINVLGGVNNILPQGRPPVPQQPVIFL	594
544	QY	GIDCVHDMTAG---RRSTAGVAGINEGMRWTFSRCTFODRGQELVDGLKVCIAQALRAW	600
595	Db	GADVTHP-PADGKKPSIAAVVGSMDAHPNRYCATVRVQQHREIITQDLAAVVBRELIQF	653
601	QY	NSCENYMSRIIVTRDVGDGQKTLVNYEVPQFLDCLKISGRGYNPRLTVIUVKRGVNT	660
654	Db	YKSTRFKPTRIIFYRDGVSEGGFQQVLHLLAIRACIKLERDYQFGITFIVVQGRHHT	713
661	QY	RPFA---QSGRLQNPLEGVIDEVTREPWDFFTVSQAVRSGSVSPTHYNVINDSG	716
714	Db	RLFCCTDKNERVKSGNIPAGTIVDTKTIHTETDFYLCSHAGTQGTSRPSHHVLMDDNR	773
717	QY	LKPDHIQRLTYKLCHIIYNNWFGVIRVPACQYAHKLAF	754
774	Db	FSSDELOIITVOLCHTVYRCRTSVSIPAPAYYAAHLVAF	811

Search completed: December 4, 2003, 19:14:08
Job time : 50 secs

GenCore version 5.1.6
Copyright (C) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model
Run on: December 6, 2003, 15:52:24 ; Search time 8508 Seconds
(without alignments)
11193.889 Million cell updates/sec

Title: US-10-043-774B-1
Perfect score: 2328
Sequence: 1 agagcttgggtggaacac.....accgcctttactacctctaa 2328

Scoring table: OLIGO_NUC
Gapop 60.0 , Gapext 60.0
Searched: 2888711 seqs, 2045481386 residues

Word size : 0
Total number of hits satisfying chosen parameters: 5777422

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database : GenEmbl.*

1: gb_ba.*
2: gb_htg.*
3: gb_in.*
4: gb_ov.*
5: gb_ov.*
6: gb_pat.*
7: gb_ph.*
8: gb_pl.*
9: gb_pr.*
10: gb_ro.*
11: gb_sts.*
12: gb_sy.*
13: gb_un.*
14: gb_vi.*
15: em_ba.*
16: em_fun.*
17: em_hum.*
18: em_in.*
19: em_mu.*
20: em_on.*
21: em_or.*
22: em_ov.*
23: em_pat.*
24: em_ph.*
25: em_pl.*
26: em_ro.*
27: em_sts.*
28: em_un.*
29: em_vi.*
30: em_htg_hum.*
31: em_htg_inv.*
32: em_htg_other.*
33: em_htg_mus.*
34: em_htg_pln.*
35: em_htg_rod.*
36: em_htg_mam.*
37: em_htg_vrt.*
38: em_sy.*
39: em_htgo_hum.*
40: em_htgo_mus.*
41: em_htgo_other.*

Pred. No. is the number of results predicted by chance to have a

score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	DB ID	Description
1	2328	100.0	2328	9	AF264004 Homo sapi
2	2246	96.5	3421	9	AF387507 Homo sapi
3	2144	92.1	3362	9	AF104260 Homo sapi
4	2144	92.1	3399	9	BC028581 Homo sapi
5	2130	91.5	2579	9	AK093133 Homo sapi
6	2022	9.5	63578	9	AC127071 Homo sapi
7	2022	8.8	193697	2	AC025837 Homo sapi
8	201	8.6	61995	2	AC090565 Homo sapi
9	129	5.5	66611	2	AC090147 Homo sapi
10	125	5.4	66611	2	AC090147 Homo sapi
11	92	4.0	461	6	BD114143 EST and e
12	58	2.5	460	6	AX209866 Sequence
13	35	1.5	3910	10	AB032604 Mus muscu
14	35	1.5	4064	10	AF438405 Mus muscu
15	34	1.5	501	6	AX395819 Sequence
16	32	1.4	39365	2	AC100141 Mus muscu
17	32	1.4	186252	2	AC116715 Mus muscu
18	32	1.4	211319	2	AC111089 Mus muscu
19	32	1.4	259213	2	AC096050 Rattus no
20	26	1.1	39365	2	AC100141 Mus muscu
21	26	1.1	61995	2	AC090565 Homo sapi
22	26	1.1	63578	9	AC127071 Homo sapi
23	26	1.1	193697	2	AC025837 Homo sapi
24	24	1.0	140348	9	AL390725 Human DNA
25	23	1.0	198389	2	AC118198 Mus muscu
26	22	0.9	89501	8	AP006420 Lotus jap
27	22	0.9	135586	9	AC073050 Homo sapi
28	22	0.9	158170	9	AC016757 Homo sapi
29	21	0.9	22	6	AX513354 Sequence
30	21	0.9	1690	3	AK115461 Ciona int
31	21	0.9	9953	1	U67553 Methanococc
32	21	0.9	40907	9	AP000358 Homo sapi
33	21	0.9	50612	9	AL513530 Human DNA
34	21	0.9	81696	9	AF438327_3 Continuation (4 of
35	21	0.9	91748	9	AC004856 Homo sapi
36	21	0.9	101507	9	AP000359 Homo sapi
37	21	0.9	110000	6	AR271569_04 Continuation (5 of
38	21	0.9	110000	9	AF438327_2 Continuation (3 of
39	21	0.9	134058	8	AC037425 Oryza sat
40	21	0.9	159139	2	AC015775 Homo sapi
41	21	0.9	162170	2	AC131452 Strongylo
42	21	0.9	177257	2	AC137121 Mus muscu
43	21	0.9	178976	2	AC118852 Rattus no
44	21	0.9	182770	9	AC104082 Homo sapi
45	21	0.9	185121	2	AC079806 Homo sapi

ALIGNMENTS

RESULT 1
AF264004
LOCUS
DEFINITION Homo sapiens HIWI mRNA, complete cds.
ACCESSION AF264004
VERSION AF264004.1 GI:15216446
KEYWORDS
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 2328)
AUTHORS Sharma,A.K., Nelson,M.C., Brandt,J.E., Wessman,M., Mahmud,N.,
Weller,K.P. and Hoffman,R.
TITLE Human CD34+ stem cells express the hiwi gene, a human homologue of

Db	1561	CAGATGAATCGAAGATGGAGAGAGCTCTGAGGTTGGACATCCCTGAGTCTGTG	1620
Qy	1621	ATGATCGTTGGCATGATGTTTACCATGACATGACAGCTGGCGGAGGTCAATCGCAGGA	1680
Db	1621	ATGATCGTTGGCATGATGTTTACCATGACATGACAGCTGGCGGAGGTCAATCGCAGGA	1680
Qy	1681	TTTGTTCGACGATCAATGAAGGATGACCGCTGGTCTCAGCTGCGATATTCAGGAT	1740
Db	1681	TTTGTTCGACGATCAATGAAGGATGACCGCTGGTCTCAGCTGCGATATTCAGGAT	1740
Qy	1741	AGAGACAGAGCTGTAGATGGCTCAAGTCTGCTCAAGCGGCTCTGAGGGCTTGG	1800
Db	1741	AGAGACAGAGCTGTAGATGGCTCAAGTCTGCTCAAGCGGCTCTGAGGGCTTGG	1800
Qy	1801	AATAGTGCATGAGTACATGCCAGCGGATCATCGTGTACCGCATGGCTAGAGAC	1860
Db	1801	AATAGTGCATGAGTACATGCCAGCGGATCATCGTGTACCGCATGGCTAGAGAC	1860
Qy	1861	GCCAGCTGAAACACCTGTGAAGTACAGTGGCAGCTTTTGTGATTTCTAAATCC	1920
Db	1861	GCCAGCTGAAACACCTGTGAAGTACAGTGGCAGCTTTTGTGATTTCTAAATCC	1920
Qy	1921	ATTGTAGAGTTTACCACTGACTAAACGTAATTTGTGTGAAGAAAGTGAACACC	1980
Db	1921	ATTGTAGAGTTTACCACTGACTAAACGTAATTTGTGTGAAGAAAGTGAACACC	1980
Qy	1981	AGATTTTCTCAGTCTGAGGAGAGCTTCAAGATCCACTTCTGAGACAGTATTGAT	2040
Db	1981	AGATTTTCTCAGTCTGAGGAGAGCTTCAAGATCCACTTCTGAGACAGTATTGAT	2040
Qy	2041	GTAGAGTTACAGACAGATGGTATGATCTTTTATCGTGAGCGGCTGTGAGAGT	2100
Db	2041	GTAGAGTTACAGACAGATGGTATGATCTTTTATCGTGAGCGGCTGTGAGAGT	2100
Qy	2101	GTTAGTGTCTCCACACATTAACATGTCATCTATGACACAGCGGCTTGAAGCCAGAC	2160
Db	2101	GTTAGTGTCTCCACACATTAACATGTCATCTATGACACAGCGGCTTGAAGCCAGAC	2160
Qy	2161	CACATACAGCGTTGACCTCAAGCTGTGCCACATCTATTACAACCTGGCCAGGTGTCAT	2220
Db	2161	CACATACAGCGTTGACCTCAAGCTGTGCCACATCTATTACAACCTGGCCAGGTGTCAT	2220
Qy	2221	CGTGTCTCTGCTCTGCGAGTACGCCCAAGCTGGCTTTCTTGTGCGCAGAGTATT	2280
Db	2221	CGTGTCTCTGCTCTGCGAGTACGCCCAAGCTGGCTTTCTTGTGCGCAGAGTATT	2280
Qy	2281	CACAGAGCGCAATCTGTCACTGTCAAAACCGCTTTTACTACCTCTAA	2328
Db	2281	CACAGAGCGCAATCTGTCACTGTCAAAACCGCTTTTACTACCTCTAA	2328
RESULT 2			
LOCUS	AF387507	3421 bp	mRNA linear PRI 01-JUL-2001
DEFINITION	Homo sapiens PIWI protein mRNA, complete cds.		
ACCESSION	AF387507		
VERSION	AF387507.1 GI:14579644		
KEYWORDS			
SOURCE	Homo sapiens (human)		
ORGANISM	Homo sapiens		
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;		
	Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.		
	1 (bases 1 to 3421)		
	Sha, J.H.		
JOURNAL	Cloning and identification of human piwi protein related to testis		
	development		
	Unpublished		
	2 (bases 1 to 3421)		
AUTHORS	Sha, J.H., Li, J.M. and Zhou, Z.M.		
	Direct Submission		
	Submitted (30-MAY-2001) Key Lab of Reproductive Medicine, Nanjing		
	Medical University, 140 Han Zhong Road, Nanjing Medical University,		

Jiangsu 210029, China	
FEATURES	Location/Qualifiers
1..3421	/organism="Homo sapiens"
	/mol_type="mRNA"
	/db_xref="taxon:9606"
	/tissue_type="testis"
	/dev_stage="adult"
CDS	64..2649
	/note="higher expression in adult testis than 6 month old embryo testis"
	/codon_start=1
	/product="PIWI protein"
	/protein_id="AAK69348.1"
	/db_xref="GI:14579645"
	/translation="MTGARARARGARQOETALVVGSTASQOQYVIQRPQPAPPAEG ELFGRGRORGTAGGTAGTQISAGFOELSLAERGRRRDHDLGVNTRONLDHVKE SKTSSGGLIVRLSTNHFRLTSPQWALQYHIDYNPLMEARLRSLALLFOHEDLLGKC HAFDGTILFLPKRLQKQVTEVSKTRNGEDVRITITLNEPFTSTCLQFNIIIFRR LUKIMNLQOIGRNYNPNPDIDIPSHRLVWPGFTTSIIQYENSIMCLTDVSHKVLRS ETVLDPMFNHYOTBEHFKQVSKELIQLVLTLYNNKTVRVDDIDMDQNPSTFK ADGSEVFLYRYKQYNGEITDLKQPLVVSOPKRERGGTLPGLPAMLIPELCYLTGL TDKMDNFNMKDLAVHRLTPEQOREVRGLLDIYHKNDVORELRDWSGLSDFSNLL SFSGRLQTEKHOGKTFDYNPQADWSKETRGAPLISVKPLDNWLLIYTERNBA NSLIQNLKFTVPAMGMOMKAIMTEVDDREAVLRVLOOKVTADTQIVVCLLSNRK KYDAIKYLCTDPTSCQCVARTLKQOQVMAIATKIALQMNKRGGLVMDVPLK LVMIYIGIDYHDMTAGRSIAGFVASINEGTRWFSRCIQDRGQELVGLVGLQAA LRWNSCNEYMPRLIIVVDGVDGQGLKLVNVEVDFDLCLKISGRGNPLRTIVY KKRVNTRFPAQGGRLQNLPGTVIDEVTREPWYDFETVSOAVRSGSVSPTHNYIVY DNSGLKPHDIQRLTYKLCHYIYNWPGVIRVPAPCQVAHLKLAFLVGOSIHRPEPNLSLN RLYYL"
	polyA signal 3354..3359
BASE COUNT	1071 a 661 c 767 g 922 t
ORIGIN	
Query Match	96.5%; Score 2246; DB 9; Length 3421;
Best Local Similarity	100.0%; Pred. No. 0;
Matches 2296;	Conservative 0; Mismatches 1; Indels 0; Gaps 0;
QY	32 TAGACCATGTTAAAGAAATCAAAAACAGGTTCTTCAGGCATTATAGTAAGTTAAGCACTA 91
DB	353 TAGACCATGTTAAAGAAATCAAAAACAGGTTCTTCAGGCATTATAGTAAGTTAAGCACTA 412
QY	92 ACCATTTCCGGCTGACATCCCGTCCCGAGTGGGCCCTTATATCAGTATCACTTACCTATA 151
DB	413 ACCATTTCCGGCTGACATCCCGTCCCGAGTGGGCCCTTATATCAGTATCACTTACCTATA 472
QY	152 ACCCACTGATGGAAGCCAGAGACTCCGTTACGCTCTCTTTTCAACAGCAAGATCTAA 211
DB	473 ACCCACTGATGGAAGCCAGAGACTCCGTTACGCTCTCTTTTCAACAGCAAGATCTAA 532
QY	212 TTGGAAGTGTCTGCTTTTGTGGAACGATATATTTTACCTTAAAGACTACAGCAAA 271
DB	533 TTGGAAGTGTCTGCTTTTGTGGAACGATATATTTTACCTTAAAGACTACAGCAAA 592
QY	272 AGGTTACTGAAGTTTATAGTAAGCCCGGAATGGAGAGATGTGAGGATAACCATCACTT 331
DB	593 AGGTTACTGAAGTTTATAGTAAGCCCGGAATGGAGAGATGTGAGGATAACCATCACTT 652
QY	332 TAACTAATGAACCTTCCACCTTACATCCCAACTTGTGTTGAGTCTCTATAATATTTTCA 391
DB	653 TAACTAATGAACCTTCCACCTTACATCCCAACTTGTGTTGAGTCTCTATAATATTTTCA 712
QY	392 GGAGGCTTTGAAATCATGAATTTGCAACAAATTTGAGCAAAATTTATTAACCCAAATG 451
DB	713 GGAGGCTTTGAAATCATGAATTTGCAACAAATTTGAGCAAAATTTATTAACCCAAATG 772
QY	452 ACCCAATGATATTTCCAAAGTCAAGTGTGTGATTTGGCTGGCTTCACTACTTCCATCC 511
DB	773 ACCCAATGATATTTCCAAAGTCAAGTGTGTGATTTGGCTGGCTTCACTACTTCCATCC 832
QY	512 TTCAAGTATGAAACAGCATCATGCTTGTGCACTGAGCTTGTAGCCATTAAGTCTCTGAGAGTG 571

MEDLINE 22033724
PUBMED 12037681
3 (bases 1 to 3362)
REFERENCE
AUTHORS Lin.H.
TITLE Direct Submission
JOURNAL Submitted (04-NOV-1998) Cell Biology, Duke University Medical Center, 412 Nanaline Duke Bldg., Research Dr., Durham, NC 27710, USA
4 (bases 1 to 3362)
REFERENCE
AUTHORS Qiao,D., Zeeman,A.-M., Deng,W., Looijenga,L.H.J. and Lin,H.
TITLE Direct Submission
JOURNAL Submitted (09-JAN-2002) Cell Biology, Duke University, 412 Nanaline Duke Building, DUMC, Durham, NC 27710, USA
REMARK Nucleotide sequence updated by submitter
COMMENT On Jan 9, 2002 this sequence version replaced gi:4038412.
FEATURES
Location/Qualifiers
1..3362
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/chromosome="12"
/map="12q24.33"
/tissue_type="testis"
57..2642
/note="member of the PIWI family"
/codon_start=1
/product="HIWI"
/protein_id="AAC97371.2"
/db_xref="GI:11809858"
/translations="MTGRARARARGRARGOETAQLVGSTASQOQGYIQRPQPPAEG
ELPGRGRORAGTAGTAKSLOGLSAGFOELSLAERGRRRDPHDLGVNTRQNDLHVKE
SKTSGSLIVRLSTNHPLRSQWALVYHIDYNPLMEARRLRSALLFOHDLIGKC
HAFDPTILFLPKLQKVTEVPSKTRNGEDVRITLITNELPTSPCTQCFVNIIFR
LLKIMLQQIGRNYIPNDIPIDPSHLVLPWPGFTSILQENSIMCLDVSXKVRIS
ETVLDFMNFYHTEHFEQESKELIGLWLTKNIKYRVDVDDIDWQNPSTFKK
ADGGSVFLYVYKQVNOETDLKQPVLVSPQRKRRPGGTLFPGPAMLIPELCYLGL
TDKRNDFNMKVLAVHETLDPQVLPQKREVRGLIDYTHKNDVQRELRLDWGLSPDSNLL
SFSGRILQTEKHGGKQTPYNPQPADWSKETRGAPLISVKPLDMLLIIVTRNVEAA
NSILQNLFLTPANGQMRVAMLEVDDETEAVRLVLOQKVADTQIVLCLSSNPKD
KIDAIKYLCTDPTSCQVARTLQKQVMAIATKIALQMKMGGEHWRVDIELK
LVNIVGDYHDMTAGRSIIVYRQVGVGDLKLVNTEVQFLDKLSIKRGVNPRLTIVV
KRWVTRFFAQSGLRQNPVGTVIDVEVTRPEWYDFIVSOAVRSGSVSPHNVIV
DNLGLPDHTLQRLTYKLCHIYNWPGVIRVPAQVYAHKLAFLVGQSIHREPNLSLN
RLYL"
BASE COUNT 1026 a 659 c 759 g 918 t
ORIGIN
Query Match 92.1%; Score 2144; DB 9; Length 3362;
Best Local Similarity 99.9%; Pred. No. 0;
Matches 2294; Conservative 0; Mismatches 3; Indels 0; Gaps 0;
QY 32 TAGACCATGTTAAGAAATCAAAAACAGGTTCTTCAGGCATTATAGTAAGGTTAAGCACTA 91
Db 346 TAGACCATGTTAAGAAATCAAAAACAGGTTCTTCAGGCATTATAGTAAGGTTAAGCACTA 405
QY 92 ACCATTTCCGGCTGACATCCCGTCCCGAGTGGCGCTTATATCAGTATCAGATCACTATA 151
Db 406 ACCATTTCCGGCTGACATCCCGTCCCGAGTGGCGCTTATATCAGTATCAGATCACTATA 465
QY 152 ACCCACTGATGGAAGCCAGAGAGATCCGGTTCAGGCTCTCTTTTCAACACAGATCTAA 211
Db 466 ACCCACTGATGGAAGCCAGAGAGATCCGGTTCAGGCTCTCTTTTCAACACAGATCTAA 525
QY 212 TTGGAAGTGTGATGCTTTTATGGAAGAGATATTTATTTTACCTAAAGACTACAGCAA 271
Db 526 TTGGAAGTGTGATGCTTTTATGGAAGAGATATTTATTTTACCTAAAGACTACAGCAA 585
QY 272 AGGTTACTGAAGTTTATGTAAGACCCGGAATGGAGGATGTGAGGATACGATCACTT 331
Db 586 AGGTTACTGAAGTTTATGTAAGACCCGGAATGGAGGATGTGAGGATACGATCACTT 645
QY 332 TAACAAATGAACCTCCACCTACATCAACCACTTGTTCAGCTTCTATATATTTTCA 391

Db 646 TAACAAATGAACCTCCACCTACATCAACCACTTGTTCAGCTTCTATATATTTTCA 705
QY 392 GGAGGCTTTTGAATCATGAATTTGCACAAATTTGGACGAATTTATTAACCCAAATG 451
Db 706 GGAGGCTTTTGAATCATGAATTTGCACAAATTTGGACGAATTTATTAACCCAAATG 765
QY 452 ACCCAATTTGATATTTCAAGTACAGGTTGGCTGATTTGGCTGGCTTCACTATCTCCATCC 511
Db 766 ACCCAATTTGATATTTCAAGTACAGGTTGGCTGATTTGGCTGGCTTCACTATCTCCATCC 825
QY 512 TTCAATGTAAGAAACAGCATCATGCTCTGCATCGATGAGCTTACCATTAAGTCTCTTCGAAGTG 571
Db 826 TTCAATGTAAGAAACAGCATCATGCTCTGCATCGATGAGCTTACCATTAAGTCTCTTCGAAGTG 885
QY 572 AGACTGTTTGGATTTCACTTTTCACTTTTATCATCAGACAGAAACATAAATTTCAAG 631
Db 886 AGACTGTTTGGATTTCACTTTTCACTTTTATCATCAGACAGAAACATAAATTTCAAG 945
QY 632 AACAAAGTTTCCAAAGAACTAATAGGTTTGTCTTACCAAGTATTAACAATAAGACAT 691
Db 946 AACAAAGTTTCCAAAGAACTAATAGGTTTGTCTTACCAAGTATTAACAATAAGACAT 1005
QY 692 ACAGAGTGGATGATATTTGATGAGTGGACCCAGAAATCCCAAGAGCCCTTTAAGAAAGCCGAG 751
Db 1006 ACAGAGTGGATGATATTTGATGAGTGGACCCAGAAATCCCAAGAGCCCTTTAAGAAAGCCGAG 1065
QY 752 GCTCTGAAGTCACTCTTCTAGATTAATCAGAGCAATTAACCAAGAGATCAACCGACT 811
Db 1066 GCTCTGGGTGACGTTCTTTAGAAATTAATCAGAGCAATTAACCAAGAGATCAACCGACT 1125
QY 812 TGAAGCAGGCTGTCTGGTTCAGCAGCCCAAGAGAGCGGGCCCTGGGGGACACTGC 871
Db 1126 TGAAGCAGGCTGTCTGGTTCAGCAGCCCAAGAGAGCGGGCCCTGGGGGACACTGC 1185
QY 872 CAGGGCTCCCATGCTCAITTCCTGAGCTCTGCTATCTTACAGGCTTAATGATAAATGC 1245
Db 932 GTAAATGATTTAACTGATGAAGAGCTTACCGGTTTATCAACAGACTTAATCCAGAGCAA 991
QY 1246 GTAAATGATTTAACTGATGAAGAGCTTACCGGTTTATCAACAGACTTAATCCAGAGCAA 1305
QY 992 GGCAGCGTGAAGTGGACGACTCATTTGATTAATTAATAAAGCAATTAATGTTCAAAGG 1051
Db 1306 GGCAGCGTGAAGTGGACGACTCATTTGATTAATTAATAAAGCAATTAATGTTCAAAGG 1365
QY 1052 AGTTTCAGAGCTGGGGTTTGAAGCTTTGATTTCAATCTTCTCTCTCAGGAAGATTT 1111
Db 1366 AGTTTCAGAGCTGGGGTTTGAAGCTTTGATTTCAATCTTCTCTCTCAGGAAGATTT 1425
QY 1112 TGCAAAACAGAAAGATTCACAGGTGGAAACATTTGATTAATTAATAAAGCAATTAATGTTCAAAGG 1171
Db 1426 TGCAAAACAGAAAGATTCACAGGTGGAAACATTTGATTAATTAATAAAGCAATTAATGTTCAAAGG 1485
QY 1172 ATTGGTCCAAAGAAACAGAGGTGCACCAATTAATTAAGTGTAAAGCCACTAGATACTGGC 1231
Db 1486 ATTGGTCCAAAGAAACAGAGGTGCACCAATTAATTAAGTGTAAAGCCACTAGATACTGGC 1545
QY 1232 TGTGTATCTATACCGGAAGAAATTTGAAGCAGCCCAATTCATTTGATTAATAAATCTATTTA 1291
Db 1546 TGTGTATCTATACCGGAAGAAATTTGAAGCAGCCCAATTCATTTGATTAATAAATCTATTTA 1605
QY 1292 AGTTTACACAGCCATGGGCATGCAATTAAGAAAGCAATTAATTAAGTGTAAAGTGAAGTGA 1351
Db 1606 AGTTTACACAGCCATGGGCATGCAATTAAGAAAGCAATTAATTAAGTGTAAAGTGAAGTGA 1665
QY 1352 GAATGAAGCTACTTTAAAGTCTTACAGCAAAAGGTTCAGCAGACACCCAGATAGTTG 1411
Db 1666 GAATGAAGCTACTTTAAAGTCTTACAGCAAAAGGTTCAGCAGACACCCAGATAGTTG 1725
QY 1412 TCTGTCTGTGTCAAGTAATTCGGAAGCAAAATACGATGCTATTATAAATAATCTGTGTA 1471

423 ACCATTCCGGCTGACATCCCGTCCCGAGTGGCGCTTATATCAGTATCATCAATGACTATA 482
QY
152 ACCACTGATGGAAGCCAGAGACTCGTTACGCTCTTCTTTTCAACAGAGACTCTAA 211
Db
483 ACCCACTGATGGAAGCCAGAGACTCGTTACGCTCTTCTTTTCAACAGAGACTCTAA 542
QY
212 TTGGAAAGTGTCTATGCTTTTGTGGAAGACGATATTTATTTTACCTTAAAGACTACAGCAA 271
Db
543 TTGGAAAGTGTCTATGCTTTTGTGGAAGACGATATTTATTTTACCTTAAAGACTACAGCAA 602
QY
272 AGTTACTGAAGTTTTPAGTAAAGCCCGAATGGAGAGATGTGAGGATTAACGATCACTT 331
Db
603 AGTTACTGAAGTTTTPAGTAAAGCCCGAATGGAGAGATGTGAGGATTAACGATCACTT 662
QY
332 TAACAAATGAATTCACCTTACATCAACCACTTTGTTGCAAGTCTTATTAATTTTCA 391
Db
663 TAACAAATGAATTCACCTTACATCAACCACTTTGTTGCAAGTCTTATTAATTTTCA 722
QY
392 GGAGGCTTTTGAATAATCATGAATTTGCAACAAATTTGACGAAATTTATTAAGCCCAATG 451
Db
723 GGAGGCTTTTGAATAATCATGAATTTGCAACAAATTTGACGAAATTTATTAAGCCCAATG 782
QY
452 ACCCAATGATATTTCCAAGTCAAGTTGTTGATTTTGGCTTGGCTTCACTACTTCCATCC 511
Db
783 ACCCAATGATATTTCCAAGTCAAGTTGTTGATTTTGGCTTGGCTTCACTACTTCCATCC 842
QY
512 TTCAGTATGAACAGCATCATGCTCTGCACTGACGTTAGCCATAAAGTCTTTCGAGTG 571
Db
843 TTCAGTATGAACAGCATCATGCTCTGCACTGACGTTAGCCATAAAGTCTTTCGAGTG 902
QY
572 AGACTGTTTGGATTTTCATGTTTCAACTTTTATCATCAGAGAGAACATATAATTTCAAG 631
Db
903 AGACTGTTTGGATTTTCATGTTTCACTTTTATCATCAGAGAGAACATATAATTTCAAG 962
QY
632 AACAGTTTCCAAGAACTAATAGTTTGTGTTTCTTACCAAGTATTAACATTAAGACAT 691
Db
963 AACAGTTTCCAAGAACTAATAGTTTGTGTTTCTTACCAAGTATTAACATTAAGACAT 1022
QY
692 ACAGAGTGGATGATTTGACTGGAGCAGCAATCCAGAGACACCTTTAAGAAAGCCGAG 751
Db
1023 ACAGAGTGGATGATTTGACTGGAGCAGCAATCCAGAGACACCTTTAAGAAAGCCGAG 1082
QY
752 GCTCTGAAGTCACTTCTTGAATCTACAGAGAGCAATAACAACAGAGATCAACCGACT 811
Db
1083 GCTCTGAAGTCACTTCTTGAATCTACAGAGAGCAATAACAACAGAGATCAACCGACT 1142
QY
812 TGAAGAGCCTGTCTTGTCTAGCCAGCCCAAGAGAGCGGGCCCTGGGGGACACTGC 871
Db
1143 TGAAGAGCCTGTCTTGTCTAGCCAGCCCAAGAGAGCGGGCCCTGGGGGACACTGC 1202
QY
872 CAGGCTTGGCATGCTCATTTCTGAGCTCTGCTATCTTACAGTCTTACATTAATATGC 931
Db
1203 CAGGCTTGGCATGCTCATTTCTGAGCTCTGCTATCTTACAGTCTTACATTAATATGC 1262
QY
932 GTAATGATTTTAAAGTGAAGAAAGCTTACAGGCTTACACAGACTTAACTCCAGAGCAA 991
Db
1263 GTAATGATTTTAAAGTGAAGAAAGCTTACAGGCTTACACAGACTTAACTCCAGAGCAA 1322
QY
992 GGCAGCGTGAAGTGGGAGCACTCATTTGATTAATCATTAATAAAGCAATATTTCAAGGG 1051
Db
1323 GGCAGCGTGAAGTGGGAGCACTCATTTGATTAATCATTAATAAAGCAATATTTCAAGGG 1382
QY
1052 AGCTTCGAGACTGGGTTTGAAGTTCGAGTCTGCTTCTTACAGAGCAATTT 1111
Db
1383 AGCTTCGAGACTGGGTTTGAAGTTCGAGTCTGCTTCTTACAGAGCAATTT 1442
QY
1112 TGCACACAGAAAGATTCACCAAGTGGAAAAACATTTGATTACAAATCCAAATTTGAG 1171
Db
1443 TGCACACAGAAAGATTCACCAAGTGGAAAAACATTTGATTACAAATCCAAATTTGAG 1502
QY
1172 ATTGGTCCAAAGAAACAGAGGTGACCAATTAATTAATTTAGTGTTAAGCCACTAGATACTGGC 1231
Db
1503 ATTGGTCCAAAGAAACAGAGGTGACCAATTAATTAATTTAGTGTTAATCCCACTAGATACTGGC 1562

1232 TGTGATCTATACGGAAGAAATTTATGAAGCAGCCAAATTTATTGATACAAAATCTATTTA 1291
QY
1563 TGTGATCTATACGGAAGAAATTTATGAAGCAGCCAAATTTATTGATACAAAATCTATTTA 1622
Db
1292 AAGTTACACAGCCATGGGCATGCAAAATGAGAAAAGCAATAATGATTGAAGTGGATGACA 1351
QY
1623 AAGTTACACAGCCATGGGCATGCAAAATGAGAAAAGCAATAATGATTGAAGTGGATGACA 1682
Db
1352 GAATGAGCCCTACTTAAAGTCTTACAGCAAAAGTTCACAGCAGACACCCAGATAGTTG 1411
QY
1683 GAATGAGCCCTACTTAAAGTCTTACAGCAAAAGTTCACAGCAGACACCCAGATAGTTG 1742
Db
1412 TCTGCTGTGTGCAAGTAATCGAAGGCAAAATACCATGCTATTAAAAAATACCTGTGTA 1471
QY
1743 TCTGCTGTGTGCAAGTAATCGAAGGCAAAATACCATGCTATTAAAAAATACCTGTGTA 1802
Db
1472 CAGATTGCCCTACCCCAAGTCAAGTGTGTGGTGGCCGGAACCTTAGGCAACAGCAAACTG 1531
QY
1803 CAGATTGCCCTACCCCAAGTCAAGTGTGTGGTGGCCGGAACCTTAGGCAACAGCAAACTG 1862
Db
1532 TCAATGCCCTTGTCTACAAAGATTCCTTACAGATGAATGCAAGATGGGAGGAGAGTCT 1591
QY
1863 TCAATGCCCTTGTCTACAAAGATTCCTTACAGATGAATGCAAGATGGGAGGAGAGTCT 1922
Db
1592 GGAGGCTGACATCCCTTGAAGTCTGATGATCGTTGGCATCGATTGTTACCATGACA 1651
QY
1923 GGAGGCTGACATCCCTTGAAGTCTGATGATCGTTGGCATCGATTGTTACCATGACA 1982
Db
1652 TGACAGCTGGGGAGGCTCAATCGAGGATTTGTTGCCAGCATCAATGAAGGAGTGAACC 1711
QY
1983 TGACAGCTGGGGAGGCTCAATCGAGGATTTGTTGCCAGCATCAATGAAGGAGTGAACC 2042
Db
1712 GCTGGTTCTCAGCTGCATATTTTACAGATAGAGACAGGAGCTGTAGATGGGCTCAAAG 1771
QY
2043 GCTGGTTCTCAGCTGCATATTTTACAGATAGAGACAGGAGCTGTAGATGGGCTCAAAG 2102
Db
1772 TCTGCTGCAAGCGGCTCTGAGGCTTGGATAGCTGCAATGAGTACATGCCAGCCGGA 1831
QY
2103 TCTGCTGCAAGCGGCTCTGAGGCTTGGATAGCTGCAATGAGTACATGCCAGCCGGA 2162
Db
1832 TCAATGCTGTAACCGCATGGCTAGGAGACGGCCAGCTGAAAAACACTGGTGAACCTACGAAG 1891
QY
2163 TCAATGCTGTAACCGCATGGCTAGGAGACGGCCAGCTGAAAAACACTGGTGAACCTACGAAG 2222
Db
1892 TGCCACAGTTTGTGATTTGTTAAATTCATTTGGTAGAGGTTACAAACCTAGACTAACGG 1951
QY
2223 TGCCACAGTTTGTGATTTGTTAAATTCATTTGGTAGAGGTTACAAACCTAGACTAACGG 2282
Db
1952 TAATTTGGTGAAGAAAGAGTGAACACCAAGATTTTGTCTGCTGAGGAGAACTTC 2011
QY
2283 TAATTTGGTGAAGAAAGAGTGAACACCAAGATTTTGTCTGCTGAGGAGAACTTC 2342
Db
2012 AGAATCCACTTCTCGAACAGTTTATGATAGAGGTTACAGACCAAGATTTGGTATGACT 2071
QY
2343 AGAATCCACTTCTCGAACAGTTTATGATAGAGGTTTACAGACCAAGATTTGGTATGACT 2402
Db
2072 TTTTATTCGTGAGCCAGGCTGTGAGAGTGTGATGTTTCTCCACACATTAAGATGCA 2131
QY
2403 TTTTATTCGTGAGCCAGGCTGTGAGAGTGTGATGTTTCTCCACACATTAAGATGCA 2462
Db
2132 TCTATGACACAGCGGCTTGAAGCCAGACCATACAGCGCTTGACCTCAAGCTGTGCC 2191
QY
2463 TCTATGACACAGCGGCTTGAAGCCAGACCATACAGCGCTTGACCTCAAGCTGTGCC 2522
Db
2192 ACATCTATTAACAATGGCCAGGCTGTCAATTCGTGTTCTCTCTGCTGAGTACGCCCAACA 2251
QY
2523 ACATCTATTAACAATGGCCAGGCTGTCAATTCGTGTTCTCTCTGCTGAGTACGCCCAACA 2582
Db
2252 AGCTGGCTTTTCTGTTGGCCAGAGTATTCACAGAGAGCCAAATCTGCTACTGTCAAAACC 2311
QY
2583 AGCTGGCTTTTCTGTTGGCCAGAGTATTCACAGAGAGCCAAATCTGCTACTGTCAAAACC 2642
Db

```

QY 2312 GCCTTTACTACTCTTAA 2328
Db 2643 GCCTTTACTACTCTTAA 2659

RESULT 5
AK093133
LOCUS
DEFINITION
  Homo sapiens cDNA FLJ35814 fis, clone TEST12006083, highly similar
  to Mus musculus mRNA for MIWI (piwi).
ACCESSION
  AK093133
VERSION
  AK093133.1 GI:21751907
KEYWORDS
  Oligo capping; fis (full insert sequence).
SOURCE
  Homo sapiens (human)
ORGANISM
  Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
  Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

REFERENCE
1
  Oshima, A., Takahashi-Fujii, A., Tanase, T., Imose, N., Takeuchi, K.,
  Arita, M., Musashino, K., Yuuki, H., Hara, H., Sugiyama, T., Irie, R.,
  Otsuki, T., Sato, H., Wakamatsu, A., Ishii, S., Yamamoto, J., Isono, Y.,
  Kawai-Hio, Y., Saito, K., Nishikawa, T., Kimura, K., Yamashita, H.,
  Matsuo, K., Nakamura, Y., Sekine, M., Kikuchi, H., Kanda, K.,
  Wagatsuma, M., Murakawa, K., Kanehori, K., Sugiyama, A., Kawakami, B.,
  Suzuki, Y., Sugano, S., Nagahari, K., Masuno, Y., Negai, K. and
  Isogai, T.
  NEDO human cDNA sequencing project
  Unpublished
  2 (bases 1 to 2579)
  Isogai, T. and Yamamoto, J.
  Direct Submission
  Submitted (04-JUL-2002) Takao Isogai, FLJ Project (HRI Team); 2-6-7
  Kazusa-Kamatari, Kisarazu, Chiba 292-0812, Japan
  (E-mail: genomics@hri.co.jp, Tel: 81-438-52-3975, Fax: 81-438-52-3986)
  NEDO human cDNA sequencing project supported by Ministry of
  Economy, Trade and Industry of Japan; cDNA full insert sequencing:
  Research Association for Biotechnology (RAB); cDNA library
  construction: Helix Research Institute (HRI) (supported by Japan
  Key Technology Center etc.); 5'- & 3'-end one pass sequencing: RAB,
  HRI, and Biotechnology Center, National Institute of Technology and
  Evaluation; clone selection for full insert sequencing: HRI and
  RAB; annotation: HRI and RAB.
  Location/Qualifiers
  1..2579
    /organism="Homo sapiens"
    /mol_type="mRNA"
    /db_xref="taxon:9606"
    /clone="TEST12006083"
    /tissue_type="testis"
    /clone_lib="TEST12"
    /notes="cloning vector: pME18SFL3"
  54..2543
    /notes="unnamed protein product"
    /codon_start=1
    /protein_id="BAC04068.1"
    /db_xref="GI:21751908"
    /translation="MTGRARARARGRQGTALVGSSTASQPGYIQRPQPPARG
    ELFGRRQGTATKASQGLQISAGFQELSLAERGRDRFDHLYNTRQNLHDVKE
    SKTSGSGTIVRLSTNHPRLTSRPOWALYQYHIDYNPLMEARLRSLFQHEDLIGKC
    HAPDGTILFLPKPQQKATVEFKTRNGEDVRIITLTNLEPPTSPTCLQFYNLI PR
    LLKTMNLOQIGRNYNRPIDIPSHRLVWPGETTSILOYENSIMLCIDVSHKVLRS
    ETVDLPNFHQTEERHFGQVSKELIGLVLLVKKYNNKRYRVDNDINDONPKSTFKK
    ADGSEVFLFYKQYKQEIITDLKQPLVSPQKRRRPGGLTLPFAMLIPELCYLTGL
    TDKMRNDNFYMKQDLAVITPEQREVGELIDYIHNDNVQRELDWGLSFDLSNL
    SFGRIILQTEKHOGKTFDYNPQADWSKETGAPLISVKPLDNLWLLIYTRNRYEA
    NSLIQNLKPTPAGMOMRKAI MEVDRTAEYLRVLOKQVTDQTLVVCLLSNRQD
    KYDAIKYLCIDCTPSCQCVARTLQKQQTWALATKIALQMNCKMGSELWRVDIPLK
    LVMTVGIDYCHDWTAGRSAGFVASINEGMTWFSRCIFQDRQQLVDGLKVCLOQA
    LRWNSCNEYMPSRIIIVRDGVDGQKLTVNVEPQFLDCLKISIGRGNPRLTVIV
    KKRVTNRFPAQSGRLQNLPGTIVDEVTREPWEVDFIVSQAVERSGVSPTHNYIV
    DNSGLKPDHQLRIKLCHYYNWPVSASTC"

FEATURES
  source
    1..2579
      /organism="Homo sapiens"
      /mol_type="mRNA"
      /db_xref="taxon:9606"
      /clone="TEST12006083"
      /tissue_type="testis"
      /clone_lib="TEST12"
      /notes="cloning vector: pME18SFL3"
      54..2543
        /notes="unnamed protein product"
        /codon_start=1
        /protein_id="BAC04068.1"
        /db_xref="GI:21751908"
        /translation="MTGRARARARGRQGTALVGSSTASQPGYIQRPQPPARG
        ELFGRRQGTATKASQGLQISAGFQELSLAERGRDRFDHLYNTRQNLHDVKE
        SKTSGSGTIVRLSTNHPRLTSRPOWALYQYHIDYNPLMEARLRSLFQHEDLIGKC
        HAPDGTILFLPKPQQKATVEFKTRNGEDVRIITLTNLEPPTSPTCLQFYNLI PR
        LLKTMNLOQIGRNYNRPIDIPSHRLVWPGETTSILOYENSIMLCIDVSHKVLRS
        ETVDLPNFHQTEERHFGQVSKELIGLVLLVKKYNNKRYRVDNDINDONPKSTFKK
        ADGSEVFLFYKQYKQEIITDLKQPLVSPQKRRRPGGLTLPFAMLIPELCYLTGL
        TDKMRNDNFYMKQDLAVITPEQREVGELIDYIHNDNVQRELDWGLSFDLSNL
        SFGRIILQTEKHOGKTFDYNPQADWSKETGAPLISVKPLDNLWLLIYTRNRYEA
        NSLIQNLKPTPAGMOMRKAI MEVDRTAEYLRVLOKQVTDQTLVVCLLSNRQD
        KYDAIKYLCIDCTPSCQCVARTLQKQQTWALATKIALQMNCKMGSELWRVDIPLK
        LVMTVGIDYCHDWTAGRSAGFVASINEGMTWFSRCIFQDRQQLVDGLKVCLOQA
        LRWNSCNEYMPSRIIIVRDGVDGQKLTVNVEPQFLDCLKISIGRGNPRLTVIV
        KKRVTNRFPAQSGRLQNLPGTIVDEVTREPWEVDFIVSQAVERSGVSPTHNYIV
        DNSGLKPDHQLRIKLCHYYNWPVSASTC"

BASE COUNT
  797 a 542 c 613 g 627 t
  ORIGIN

```

```

Query Match 91.5%; Score 2130; DB 9; Length 2579;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 2180; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 32 TAGACCATGTTAAAGAAATCAAAAAACAGGTTCTTCAGGCAATTATAGTAAGGTTAAAGCACTA 91
Db 343 TAGACCATGTTAAAGAAATCAAAAAACAGGTTCTTCAGGCAATTATAGTAAGGTTAAAGCACTA 402
QY 92 ACCATTTCGGGTGCATCCCGTCCCGTCCCGTCCCGTCCCGTCCCGTCCCGTCCCGTCCCGTCCCGT 151
Db 403 ACCATTTCGGGTGCATCCCGTCCCGTCCCGTCCCGTCCCGTCCCGTCCCGTCCCGTCCCGTCCCGT 462
QY 152 ACCACTGATGAAGCAGCAGACATCCCGTCCCGTCCCGTCCCGTCCCGTCCCGTCCCGTCCCGTCCCGT 211
Db 463 ACCCACTGATGAAGCAGCAGACATCCCGTCCCGTCCCGTCCCGTCCCGTCCCGTCCCGTCCCGTCCCGT 522
QY 212 TTGGAAAGTGTCTATGCTTTTGTGATGGAACGATATATTTTACCTAAAAGACTACAGCAAA 271
Db 523 TTGGAAAGTGTCTATGCTTTTGTGATGGAACGATATATTTTACCTAAAAGACTACAGCAAA 582
QY 272 AGGTTACTGAAGTTTTTATAGTAAGCCCGGAATGGAGAGGATGTGAGGATAAGCATCACTT 331
Db 583 AGGCTACTGAAGTTTTTATAGTAAGCCCGGAATGGAGAGGATGTGAGGATAAGCATCACTT 642
QY 332 TAACAAATGAACCTCCACCTCATCATCAACCACTTTGTCAGTCTCTATATATATTTTCA 391
Db 643 TAACAAATGAACCTCCACCTCATCATCAACCACTTTGTCAGTCTCTATATATATTTTCA 702
QY 392 GGAGGCTTTTGAATAATCATGAATTTGCAACAAATGGACGAAATATTTAACCACCAATG 451
Db 703 GGAGGCTTTTGAATAATCATGAATTTGCAACAAATGGACGAAATATTTAACCACCAATG 762
QY 452 ACCCAATTCATATTCCAAGTGCACAGTTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGT 511
Db 763 ACCCAATTCATATTCCAAGTGCACAGTTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGT 822
QY 512 TTCAGTATGAACAACAGCATCATCTGTCACATGACCTAGCCATAGCTTAAAGTCCCTTCGAAGTG 571
Db 823 TTCAGTATGAACAACAGCATCATCTGTCACATGACCTAGCCATAGCTTAAAGTCCCTTCGAAGTG 882
QY 572 AGACTCTTTTGGATTTTCATGTTCAACTTTTATCATCAGACAGAAACATAAATTTCAAG 631
Db 883 AGACTCTTTTGGATTTTCATGTTCAACTTTTATCATCAGACAGAAACATAAATTTCAAG 942
QY 632 AACCAATTTCCAAAGAACTAATAGTTTGTGTTCTTCCAAAGTATACAAATAGACAT 691
Db 943 AACCAATTTCCAAAGAACTAATAGTTTGTGTTCTTCCAAAGTATACAAATAGACAT 1002
QY 692 ACAGTGTGATGATATTGACTGGGACCAAGTCCCAAGAGCACCTTTAAGAAAGCCGAGC 751
Db 1003 ACAGTGTGATGATATTGACTGGGACCAAGTCCCAAGAGCACCTTTAAGAAAGCCGAGC 1062
QY 752 GCTCTGAAGTCAAGCTTTTAGAATACCTACAGGAAGCAATACCAACAGAGATCACCGACT 811
Db 1063 GCTCTGAAGTCAAGCTTTTAGAATACCTACAGGAAGCAATACCAACAGAGATCACCGACT 1122
QY 812 TGAAGCAGCCTGCTTGGTGCAGCCAGCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 871
Db 1123 TGAAGCAGCCTGCTTGGTGCAGCCAGCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1182
QY 872 CAGGGCTCGCATGCTTCATTCCTGAGCTCTGCTATCTTACAGCTCTAACTGATAAATGC 931
Db 1183 CAGGGCTCGCATGCTTCATTCCTGAGCTCTGCTATCTTACAGCTCTAACTGATAAATGC 1242
QY 932 GTAATGATTTTAAAGTGTGAAGAGCTTAGCCGCTTCATACAGACTAACTCCAGAGCAAA 991
Db 1243 GTAATGATTTTAAAGTGTGAAGAGCTTAGCCGCTTCATACAGACTAACTCCAGAGCAAA 1302
QY 992 GGCAGGCTGAAGTGGGACGACCTATTGATTACATTAATAAAGCGATAATGTTCAAAGGG 1051
Db 1303 GGCAGGCTGAAGTGGGACGACCTATTGATTACATTAATAAAGCGATAATGTTCAAAGGG 1362

```

QY 1052 AGCTTCGAGACTGGGTTTGAGCTTTGATTCACAACTTACTGTCTCTCTCAGGAAGATTT 1111
DB 1363 AGCTTCGAGACTGGGTTTGAGCTTTGATTCACAACTTACTGTCTCTCTCAGGAAGATTT 1422
QY 1112 TGCAACACAGAAAAGATTCACCAAGGTGGAAAAACATTTGATTACATTCACCAATTTGCGAG 1171
DB 1423 TGCAACACAGAAAAGATTCACCAAGGTGGAAAAACATTTGATTACATTCACCAATTTGCGAG 1482
QY 1172 ATTGGTCCAAAGAAAACAGAGGTGCACCATTAATTAGTGTAAAGCCACTAGATACTGCG 1231
DB 1483 ATTGGTCCAAAGAAAACAGAGGTGCACCATTAATTAGTGTAAAGCCACTAGATACTGCG 1542
QY 1232 TGTGTATCTATACGCGAAGAAAATTTATGACGAGCAATTCATTGATACAAAATCTATTTA 1291
DB 1543 TGTGTATCTATACGCGAAGAAAATTTATGACGAGCAATTCATTGATACAAAATCTATTTA 1502
QY 1292 AAGTTACACAGCCATGGCGATCAATGAGAAAAGCAATTAATGATGAAGTGGATGACA 1351
DB 1603 AAGTTACACAGCCATGGCGATCAATGAGAAAAGCAATTAATGATGAAGTGGATGACA 1662
QY 1352 GAATGAGCCTACTTAAGAGTCTTACAGCAAAAGGTACAGCAGACACCCAGATAGTTG 1411
DB 1663 GAATGAGCCTACTTAAGAGTCTTACAGCAAAAGGTACAGCAGACACCCAGATAGTTG 1722
QY 1412 TCTGTCTGTGTCAAGTAATCGGAAGGACAAATACGATGCTATTAAAAATACCTGTGTA 1471
DB 1723 TCTGTCTGTGTCAAGTAATCGGAAGGACAAATACGATGCTATTAAAAATACCTGTGTA 1782
QY 1472 CAGATTGCCCTACCCCAAGTCAGTGTGTGGTGGCCGAACCTTAGGCAAAACCAAACTG 1531
DB 1783 CAGATTGCCCTACCCCAAGTCAGTGTGTGGTGGCCGAACCTTAGGCAAAACCAAACTG 1842
QY 1532 TCATGGCCATGCTACAAAGATTCGCCCTACAGATGAATCGAGATGGGAGGAGCTCT 1591
DB 1843 TCATGGCCATGCTACAAAGATTCGCCCTACAGATGAATCGAGATGGGAGGAGCTCT 1902
QY 1592 GAGGTTGGACATCCCTCGAGCTCGTGATGCTGTGGTGGATCGATTGTACCATGACA 1651
DB 1903 GAGGTTGGACATCCCTCGAGCTCGTGATGCTGTGGTGGATCGATTGTACCATGACA 1962
QY 1652 TGACAGCTGGCGGAGGTCAATCGCAGGATTTGTTGCGAGCATCAATGAAGGATGACCC 1711
DB 1963 TGACAGCTGGCGGAGGTCAATCGCAGGATTTGTTGCGAGCATCAATGAAGGATGACCC 2022
QY 1712 GCTGTTTCTCAGCTGCATATTTACAGATAGAGCAGGAGCTGGTATAGTGGCTCAAG 1771
DB 2023 GCTGTTTCTCAGCTGCATATTTACAGATAGAGCAGGAGCTGGTATAGTGGCTCAAG 2082
QY 1772 TCTGCTGCAAGCGGCTCTGAGGCTTGGATAGCTGCAATGAGTACATGCCAGCCGGA 1831
DB 2083 TCTGCTGCAAGCGGCTCTGAGGCTTGGATAGCTGCAATGAGTACATGCCAGCCGGA 2142
QY 1832 TCATCTGTACCGGATGGCGTAGGACGCGCAGCTGAAACACTGGTGAATCAGGAG 1891
DB 2143 TCATCTGTACCGGATGGCGTAGGACGCGCAGCTGAAACACTGGTGAATCAGGAG 2202
QY 1892 TGCCACAGTTTGGATGTTCTAAATCCATTTGATAGGTTTCAACCCCTAGCTAAGG 1951
DB 2203 TGCCACAGTTTGGATGTTCTAAATCCATTTGATAGGTTTCAACCCCTAGCTAAGG 2262
QY 1952 TAAATTGGTGAAGAAAAGAGTGAACACAGATTTTGTGCTGAGTGGAGGAAGCTTC 2011
DB 2263 TAAATTGGTGAAGAAAAGAGTGAACACAGATTTTGTGCTGAGTGGAGGAAGCTTC 2322
QY 2012 AGAATCCACTTCTCGGAACAGTATTGATGTAGAGGTTTACAGACCCAGAAATGGATGACT 2071
DB 2323 AGAATCCACTTCTCGGAACAGTATTGATGTAGAGGTTTACAGACCCAGAAATGGATGACT 2382
QY 2072 TTTTATCGTAGGACGCTGTGAGAGTGGTATGTTCTCCACACATTAACATGTC 2131
DB 2383 TTTTATCGTAGGACGCTGTGAGAGTGGTATGTTCTCCACACATTAACATGTC 2442
QY 2132 TCTATGACAAACGCGGCTTGAGCCAGACCATACACGCGCTTGACCTACAGCTGTGCG 2191

DB 2443 TCTATGACAAACGCGGCTTGAGCCAGACCATACAGCGCTTGACCTACAGCTGTGCG 2502
QY 2192 ACATCTATTACAACTGGCCAG 2212
DB 2503 ACATCTATTACAACTGGCCAG 2523

RESULT 6

AC127071 63578 bp DNA linear PRI 03-SEP-2002
Homo sapiens 12 BAC RP11-200K12 (Roswell Park Cancer Institute
Human BAC Library) complete sequence.

AC127071
AC127071.3 GI:22657480

HTG.

Homo sapiens (human)

Homo sapiens

ORGANISM

REFERENCE

AUTHORS

1 (bases 1 to 63578)
Muzny,D.M., Adams,C., Adio-Oduola,B., Ali-osman,F.R., Allen,C.,
Alsbrooks,S.L., Amaraturge,H.C., Are,J.R., Ayale,M., Banks,T.,
Barbaria,J., Benton,J., Bimage,K., Blankenburg,K., Bonnin,D.,
Bouck,J., Bowie,S., Brieva,M., Brown,E., Brown,M., Bryant,N.P.,
Buhay,C., Burch,P., Burkett,C., Burrell,K.L., Byrd,N.C.,
Carton,T.F., Carter,M., Cavazos,S.R., Chacko,J., Chavez,D.,
Chen,G., Chen,R., Chen,Z., Chiu,D., Chowdhry,I., Christopoulos,C.,
Cleveland,C.D., Cox,C., Coyie,M.D., Dathorne,S.R., David,R.,
David,M.L., Davis,C., Davy-Carroll,L., Dederich,D.A.,
Delaney,K.R., Delgado,C., Denn,A.L., Ding,Y., Dinh,H.H.,
Dunaway,K.J., Draper,H., Dugan-Rocha,S., Durbin,K.J.,
Earnhart,C., Edgar,D., Edwards,C.C., Elhaj,C., Emerling,S.,
Escotto,M., Falls,T., Ferraguto,D., Flagg,N., Ford,J., Foster,P.,
Frantz,P., Gabisi,A., Gao,J., Garcia,A., Garner,F., Garza,N.,
Gill,R., Gorrell,J.H., Guevara,W., Gunaratne,P., Hale,S.,
Hamilton,K., Han,J., Harris,C., Harris,K., Hart,M., Haylak,P.,
Hawes,A., Hernandez,J., Hernandez,O., Hodgson,A., Hogues,M.,
Holloway,C., Hollins,B., Honsi,F., Howard,S., Huber,J., Hulyk,S.,
Hume,J., Ioshikhes,I., Jackson,L.E., Jacobson,B., Jia,Y.,
Johnson,R., Jolivet,S., Joudah,S., Karlsson,E., Kelly,S., Khan,U.,
King,L., Korvah,J., Kovar,C., Kratovic,J., Kureshi,A., Landry,N.,
Leal,B., Lee,E., Lewis,L.C., Lewis,L., Li,J., Li,Z., Lichtarge,O.,
Lieu,C., Liu,J., Liu,W., Loulseged,H., Lozada,R.J., Lu,X.,
Lucier,A., Lucier,R., Luna,R., Ma,J., Maheshwari,M., Mapua,P.,
Marondel,I., Martin,R., Martindale,A., Martinez,E., Massey,E.,
Mawhiney,E., McLeod,M.P., Meador,M., Mei,G., Merscher,S.,
Metzker,M., Miller,A., Miner,G., Miner,Z., Mitchell,T.,
Mohabbat,K., Montgomery,K.T., Morgan,M., Morris,S., Moser,M.,
Neal,D., Nelson,D., Newton,J., Newton,N., Nguyen,A., Nguyen,N.,
Nguyen,N., Nickerson,E., Nwokenkwo,S., Oguh,M., Okwuonu,G.,
Oraguine,N., Oviedo,R., Pace,A., Payton,B., Peery,J., Perez,L.,
Peters,L., Pickens,R., Primus,E., Fu,L.L., Quiles,M., Ren,Y.,
Rives,M., Rojas,A., Rojubokan,I., Rolfe,M., Ruiz,S., Savery,G.,
Scherer,S., Scott,G., Shen,H., Shim,C., Shoostari,N., Sisson,I.,
Sodergren,E., Sonaika,T., Sparks,A., Stanley,H., Stone,H.,
Sutton,A., Svatek,A., Tabor,P., Tamerisa,A., Tamerisa,K., Tang,H.,
Tansay,J., Taylor,C., Taylor,T., Telford,B., Thomas,N., Thomas,S.,
Umani,K., Vasquez,L., Vera,V., Villalobos,B., Vinson,R., Wang,Q.,
Wang,S., Ward-Moore,S., Warren,R., Washington,C., Watlington,S.,
Williams,G., Williamson,A., Wleczky,R., Wooden,S., Worley,K.,
Wu,C., Wu,Y., Wu,Y.F., Zhou,J., Zorrilla,S., Kuchelapati,R.,
Weinstock,G., and Gibbs,R.

Direct Submission

Unpublished

2 (bases 1 to 63578)

Worley,K.C.

Direct Submission

TITLE

JOURNAL

REFERENCE

AUTHORS

JOURNAL

REFERENCE

AUTHORS

Submitted (13-JUL-2002) Human Genome Sequencing Center, Department
of Molecular and Human Genetics, Baylor College of Medicine, One
Baylor Plaza, Houston, TX 77030, USA
3 (bases 1 to 63578)
Worley,K.C.

TITLE
JOURNAL
 Submitted (01-AUG-2002) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA

REFERENCE
AUTHORS
JOURNAL
 Worley, K.C.
 Direct Submission
 Submitted (03-SEP-2002) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA

COMMENT
 On Sep 3, 2002 this sequence version replaced gi:21908383.
 INFORMATION: <http://www.hgsc.bcm.tmc.edu/> or email gc-help@bcm.tmc.edu

CLONE LENGTH: This sequence does not necessarily represent the entire insert of this clone. Overlapping regions of clones are only sequenced and submitted once, so the sequence for the remainder of the insert may be found in the record for the adjacent clones. Overlapping clones are noted at the beginning and end of the Features listing.

ANNOTATION OF FEATURES:
 STSs are identified using ePCR (Genome Res. 7:541-550) searches of a local database that includes entries from dbSTS, GDB, and local mapping efforts.
 Repeats are identified using RepeatMasker (A. Smit and P. Green, unpublished.) for Human and Mouse sequences.
 Genes and Region of sequence similarity are identified by BLAST (Nuc. Acids Res. 25:3389-3402) similarity (expect < 1e-34) to the EST and cDNA sequences. Genes demonstrate at least two exons flanked by consensus splice sites that maintained sequence continuity across the splice junctions. Sequences that are not identical matches are annotated as similar.

SEQUENCING READ COVERAGE: Sequencing is completed to a minimum standard of double strand coverage with a minimum of 2 clones and 2 reads with no ambiguities or 2 chemistries with a minimum of 2 clones and 3 reads with no ambiguities. If the sequence quality for a region does not meet this standard, it will be indicated in the annotation as Low Coverage.

QUALITY OF INDIVIDUAL BASES: This sequence meets stringent quality standards - estimated error rate less than 1 per 10,000 bases. Reports of lowest quality individual bases and measures of base quality are listed below. Description of the metrics can be found at URL:
<http://www.hgsc.bcm.tmc.edu:8088/quality.info/genbank.annotation.html>.

FEATURES
source
 Location/Qualifiers
 1..63578
 /organism="Homo sapiens"
 /mol_type="genomic DNA"
 /db_xref="taxon:9606"
 /chromosome="12"
 /clone="RP11-200K12"
 complement(1..2013)
 /note="overlaps bases 1..2013 of clone AC026764"
 /functions="clone overlap"
 1694..1985
 /standard_name="G64712"
 15975..16062
 /function="single clone coverage"
 19548..19627
 /note="sized by PCR to be 80 bps and confirmed by restriction digest"
 /function="unresolved dinucleotide repeat"
 50966..51098
 /standard_name="RH102272"
 complement(61575..63578)
 /note="overlaps bases 190771..192774 of clone AC063926"
 /function="clone overlap"
 61711..62052
 /standard_name="RH122074"

misc_feature
 complement(1..2013)
 /note="overlaps bases 1..2013 of clone AC026764"
 /functions="clone overlap"
 1694..1985
 /standard_name="G64712"
 15975..16062
 /function="single clone coverage"
 19548..19627
 /note="sized by PCR to be 80 bps and confirmed by restriction digest"
 /function="unresolved dinucleotide repeat"
 50966..51098
 /standard_name="RH102272"
 complement(61575..63578)
 /note="overlaps bases 190771..192774 of clone AC063926"
 /function="clone overlap"
 61711..62052
 /standard_name="RH122074"

BASE COUNT 16941 a 13136 c 14160 g 19341 t
ORIGIN
 Query Match 9.5%; Score 222; DB 9; Length 63578;
 Best Local Similarity 100.0%; Pred. No. 5.6e-121;
 Matches 222; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 56 CAGGTTCTTCAGGCATTATAGTAAAGTAAAGCACTAAACCATTCGGCTGACATCCCGTC 115
 Db 25305 CAGGTTCTTCAGGCATTATAGTAAAGTAAAGCACTAAACCATTCGGCTGACATCCCGTC 25364
 QY 116 CCCAGTGGCCCTTATATCATGATATCATGATATCAATGATATCAATGATGAGCCAGAGAC 175
 Db 25365 CCCAGTGGCCCTTATATCATGATATCATGATATCAATGATGAGCCAGAGAC 25424
 QY 176 TCGTTTCAGCTCTCTCTTTTCAACAGAGATCTAATTCGAAAGTGTCTATGCTTTTGATG 235
 Db 25425 TCGTTTCAGCTCTCTTTTCAACAGAGATCTAATTCGAAAGTGTCTATGCTTTTGATG 25484
 QY 236 GAACGATATATTTTACCTAAAGACTACAGCAAAAGGTTA 277
 Db 25485 GAACGATATATTTTACCTAAAGACTACAGCAAAAGGTTA 25526

RESULT 7
 AC025837 193697 bp DNA linear HTG 06-MAY-2001
 LOCUS Homo sapiens clone RP11-200K12, WORKING DRAFT SEQUENCE, 29
 DEFINITION unordered pieces.
 AC025837
 ACCESSION AC025837.3 GI:8072609
 VERSION HTG; HTGS_PHASE1; HTGS_DRAFT.
 KEYWORDS Homo sapiens (human)
 SOURCE
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 Birren,B., Linton,L., Nusbaum,C. and Lander,E.
 1 (bases 1 to 193697)
 Homo sapiens chromosome, clone RP11-200K12
 Unpublished
 2 (bases 1 to 193697)
 Birren,B., Linton,L., Nusbaum,C., Lander,E., Abraham,H., Allen,N., Anderson,S., Baldwin,J., Barna,N., Bastien,V., Bedalov,F., Boguski,M., Brown,A., Brown,A., Burkett,G., Campopiano,A., Castle,A., Choe,Y., Colangelo,M., Collins,S., Collymore,A., Cooke,P., DeArnell,K., Dewar,K., Diaz,J.S., Dodge,S., Domino,M., Doyle,M., Ferrelle,P., FitzHugh,W., Gage,D., Galagan,J., Gardyna,S., Ginde,S., Goyette,M., Graham,L., Grand-Pierre,N., Grant,G., Hagos,B., Heaford,A., Horton,L., Howland,J.C., Iliev,I., Johnson,R., Jones,C., Kann,L., Karatas,A., Klein,J., LaRocque,K., Lamazares,R., Landers,T., Lehotsky,J., Levine,R., Liu,C., Liu,G., Locke,K., Macdonald,P., Marguis,N., McCarthy,M., McEwan,P., McGurk,A., McKernan,K., McPheeters,R., Meldrum,J., Meneus,L., Mihova,T., Miranda,C., Mlenga,V., Morrow,J., Murphy,T., Naylor,J., Norman,C.H., O'Connor,T., O'Donnell,P., O'Neill,D., Oliveri,W., Oliver,J., Peterson,K., Pierre,N., Pisanic,B., Pollara,V., Raymond,C., Riley,R., Rogov,P., Rochnan,D., Roy,A., Santos,R., Schauer,S., Severy,P., Spencer,B., Stange-Thomann,N., Stojanovic,N., Subramanian,A., Talamas,J., Tesfaye,S., Theodore,J., Titrell,A., Travers,M., Trigilio,J., Vassiliev,H., Viel,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Ye,W.J., Young,G., Zainou,J., Zimmer,A. and Zody,M.
 Direct Submission
 Submitted (16-MAR-2000) Whitehead Institute/MIT Center for Genome Research, 320 Charles Street, Cambridge, MA 02141, USA
 On May 25, 2000 this sequence version replaced gi:7342026.
 All repeats were identified using RepeatMasker:
 Smit, A.F.A. & Green, P. (1996-1997)
<http://ftp.genome.washington.edu/RW/RepeatMasker.html>
 ----- Genome Center
 Center: Whitehead Institute/ MIT Center for Genome Research
 Center code: WIBR
 Web site: <http://www-seq.wi.mit.edu>

*	1	1628:	contig	of	1628	bp	in	length
*	1529	1728:	gap	of	100	bp		
*	1729	2962:	contig	of	1234	bp	in	length
*	2363	3062:	gap	of	100	bp		
*	3063	4989:	contig	of	1927	bp	in	length
*	4990	5089:	gap	of	100	bp		
*	5090	6666:	contig	of	1577	bp	in	length
*	6667	6766:	gap	of	100	bp		
*	6767	8667:	contig	of	1901	bp	in	length
*	8668	8667:	gap	of	100	bp		
*	8768	11259:	contig	of	2492	bp	in	length
*	11260	11359:	gap	of	100	bp		
*	11360	13557:	contig	of	2158	bp	in	length
*	13558	13657:	gap	of	100	bp		
*	13658	17173:	contig	of	3516	bp	in	length
*	17174	17273:	gap	of	100	bp		
*	17274	20060:	contig	of	2787	bp	in	length
*	20061	20160:	gap	of	100	bp		
*	20161	23843:	contig	of	3683	bp	in	length
*	23844	23943:	gap	of	100	bp		
*	23944	27203:	contig	of	3260	bp	in	length
*	27204	27303:	gap	of	100	bp		
*	27304	31592:	contig	of	4289	bp	in	length
*	31593	31692:	gap	of	100	bp		
*	31693	36740:	contig	of	5048	bp	in	length
*	36741	36840:	gap	of	100	bp		
*	36841	42197:	contig	of	5357	bp	in	length
*	42198	42297:	gap	of	100	bp		
*	42298	47652:	contig	of	5355	bp	in	length
*	47653	47752:	gap	of	100	bp		
*	47753	52307:	contig	of	4555	bp	in	length
*	52308	52407:	gap	of	100	bp		
*	52408	58606:	contig	of	6199	bp	in	length
*	58607	58706:	gap	of	100	bp		
*	58707	65584:	contig	of	6878	bp	in	length
*	65585	65684:	gap	of	100	bp		
*	65685	72492:	contig	of	6808	bp	in	length
*	72493	72592:	gap	of	100	bp		
*	72593	78756:	contig	of	6164	bp	in	length
*	78757	78856:	gap	of	100	bp		
*	78857	85474:	contig	of	6618	bp	in	length
*	85475	85574:	gap	of	100	bp		
*	85575	91951:	contig	of	6377	bp	in	length
*	91952	92051:	gap	of	100	bp		
*	92052	99044:	contig	of	6993	bp	in	length
*	99045	99144:	gap	of	100	bp		
*	99145	104884:	contig	of	5740	bp	in	length
*	104885	104984:	gap	of	100	bp		
*	104985	112130:	contig	of	7146	bp	in	length

FEATURES

misc_feature 126420..137651
/note="assembly fragment"
misc_feature 137752..193697
/note="assembly fragment"
BASE COUNT 55529 a 42321 c 40366 g 52678 t 2803 others
ORIGIN

Query Match 8.8%; Score 205; DB 2; Length 193697;
Best Local Similarity 100.0%; Pred. No. 1e-110;
Matches 205; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 473 ACAGTTGGTGAATGGCCCTGCTTCACTACTTCCATCTCCATCTTCACTAGTGAAGAACAGCATCA 532
DB 161209 ACAGTTGGTGAATGGCCCTGCTTCACTACTTCCATCTCCATCTTCACTAGTGAAGAACAGCATCA 161150
QY 533 TGCTCTGCAGTACGTTAGCCATAAAGTCTTCCAAAGTGAAGTCTTTGGATTTCATGT 592
DB 161149 TGCTCTGCAGTACGTTAGCCATAAAGTCTTCCAAAGTGAAGTCTTTGGATTTCATGT 161090
QY 593 TCAACTTTTATCATCAGACAGAGAATAAAATTTCAAGAACAGTTTCCAAAGACTAA 652
DB 161089 TCAACTTTTATCATCAGACAGAGAATAAAATTTCAAGAACAGTTTCCAAAGACTAA 161030
QY 653 TAGCTTAGTGTCTTACCAAGTA 677
DB 161029 TAGCTTAGTGTCTTACCAAGTA 161005

RESULT 8
AC090565 61995 bp DNA linear HTG 03-MAR-2001
LOCUS Homo sapiens chromosome 8 clone RP11-440B20 map 8, LOW-PASS
DEFINITION SEQUENCE SAMPLING.
AC090565
VERSION AC090565.1 GI:13194295
KEYWORDS HTG; HTGS-PHASE0.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Mammalia; Euthera; Primates; Catarrhini; Homiidae; Homo.
REFERENCE 1 (bases 1 to 61995)
Barnes, N., Bastien, V., Bouglavsky, L., Boukhgalter, B., Brown, A.,
Camata, J., Campopiano, A., Chospe, Y., Colangelo, M., Collins, S.,
Collamore, A., Cooke, P., DeArillano, K., Dewar, K., Diaz, J.S.,
Dodge, S., Faro, S., Ferreira, P., FitzHugh, W., Gage, D., Galagan, J.,
Gardyna, S., Ginde, S., Goyette, M., Graham, L., Grand-Pierre, N.,
Hagos, B., Heaford, A., Horton, L., Hulme, W., Iliev, I., Johnson, R.,
Jones, C., Karatas, A., LaRoque, K., Lamazares, R., Landers, T.,
Lehoczky, J., Levine, R., Liu, G., MacLean, C., Macdonald, P.,
Marquis, N., Mathews, C., McCarthy, M., McEwan, P., McKernan, K.,
McPheters, R., Meldrum, J., Meneus, L., Minova, T., Mlenga, V.,
Murphy, T., Naylor, J., Nguyen, C., Norbu, C., Norman, C.H.,
O'Connor, T., O'Donnell, P., O'Neill, D., Oliver, J., Peterson, K.,
Phunkhang, P., Pierre, N., Pollara, V., Raymond, C., Retta, R.,
Rieback, M., Riley, R., Rise, C., Rogov, P., Roman, J., Rosetti, M.,
Roy, A., Santos, R., Schauer, S., Schupback, R., Seaman, S., Severy, P.,
Sougniez, C., Spencer, B., Stange-Thomann, N., Stojanovic, N.,
Strauss, N., Subramanian, A., Talamas, J., Tesfaye, S., Theodore, J.,
Travers, M., Travis, N., Trigilio, J., Vassiliev, H., Viel, R., Vo, A.,
Wilson, B., Wu, X., Wyman, D., Ye, W.J., Young, G., Zainoun, J.,
Zembek, L., Zimmer, A. and Zody, M.

Direct Submission
TITLE Submitted (03-MAR-2001) Whitehead Institute/MIT Center for Genome
JOURNAL Research, 320 Charles Street, Cambridge, MA 02141. USA
COMMENT All repeats were identified using RepeatMasker:
Smit, A.F.A. & Green, P. (1996-1997)
http://ftp.genome.washington.edu/RM/RepeatMasker.html
----- Genome Center

Center: Whitehead Institute/ MIT Center for Genome Research
Center code: WIBR
Web site: <http://www-seq.wi.mit.edu>
Contact: sequence_submissions@genome.wi.mit.edu
----- Project Information
Center project name: L11615
Center Clone name: 440_B_20

* NOTE: This record contains 78 individual
* sequencing reads that have not been assembled into
* contigs. Runs of N are used to separate the reads
* and the order in which they appear is completely
* arbitrary. Low-pass sequence sampling is useful for
* identifying clones that may be gene-rich and allows
* overlap relationships among clones to be deduced.
* However, it should not be assumed that this clone
* will be sequenced to completion. In the event that
* the record is updated, the accession number will
* be preserved.

1
692 791: contig of 691 bp in length
792 791: gap of 100 bp
792 1505: contig of 714 bp in length
1506 1605: gap of 100 bp
1606 2290: contig of 685 bp in length
2291 2390: gap of 100 bp
2391 3097: contig of 707 bp in length
3098 3197: gap of 100 bp
3198 3898: contig of 701 bp in length
3899 4681: contig of 683 bp in length
4682 4781: gap of 100 bp
4782 5452: contig of 671 bp in length
5453 6286: contig of 734 bp in length
6287 7117: contig of 731 bp in length
7118 7217: gap of 100 bp
7218 7935: contig of 718 bp in length
7936 8035: gap of 100 bp
8036 8734: contig of 699 bp in length
8735 9537: contig of 703 bp in length
9538 9638: gap of 100 bp
9639 10318: contig of 681 bp in length
10319 10418: gap of 100 bp
10419 11120: contig of 702 bp in length
11121 11220: gap of 100 bp
11221 11900: contig of 680 bp in length
11901 12000: gap of 100 bp
12001 12700: contig of 700 bp in length
12701 12800: gap of 100 bp
12801 13493: contig of 693 bp in length
13494 14282: contig of 689 bp in length
14283 14382: gap of 100 bp
14383 15074: contig of 692 bp in length
15075 15174: gap of 100 bp
15175 15847: contig of 673 bp in length
15848 15947: gap of 100 bp
15949 16625: contig of 678 bp in length
16626 16725: gap of 100 bp
16726 17417: contig of 692 bp in length
17418 17517: gap of 100 bp
17518 18311: contig of 694 bp in length
18312 18996: contig of 685 bp in length
18997 19096: gap of 100 bp
19097 19756: contig of 660 bp in length
19757 20563: contig of 707 bp in length
20564 21383: contig of 720 bp in length
21384 21483: gap of 100 bp

* 50895	50994: gap of 100 bp
* 51680:	contig of 686 bp in length
* 51780:	gap of 100 bp
* 51781:	contig of 694 bp in length
* 52574:	gap of 100 bp
* 52575:	contig of 680 bp in length
* 53254:	gap of 100 bp
* 53354:	contig of 706 bp in length
* 53355:	gap of 100 bp
* 54160:	contig of 699 bp in length
* 54161:	gap of 100 bp
* 54860:	contig of 100 bp

8.6%; Score 201; DB 2; Length 61995;
Query Match Best Local Similarity 100.0%; Pred. No. 2.6e-108;
Matches 201; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1408	GTGTCGTCGCTGTTGCAAGTAATCGGAAGACAAATACGATGCATTATAAAAATACCTG	1467
Db		22748
QY 22689	GTTGCTGCTGTTGTCAGATGATCGAAGACAAATACGATGCATTATAAAAATACCTG	22748
Db		
QY 1468	TGTACAGATTGCCCTACCCCAAGTCAGTGTGTGGCCCGAACCTTAGGCACAAAGCAA	1527
Db	TGTACAGATTGCCCTACCCCAAGTCAGTGTGTGGCCCGAACCTTAGGCACAAAGCAA	22808
QY 1528	ACTGTCATGCGCATTCGTACAAAGATTGCGCTACAGATGAACCTCAGATGGAGGACAG	1587
Db	ACTGTCATGCGCATTCGTACAAAGATTGCGCTACAGATGAACCTCAGATGGAGGACAG	22809
QY 1588	CTCTGGAGGGTGACATCCCC	1608
Db		
QY 22869	CTCTGGAGGGTGACATCCCC	22889

RESULT 9
AC090147/c
LOCUS
DEFINITION Homo sapiens chromosome 8 clone RP11-382A18 map 8, LOW-PASS
SEQUENCE SAMPLING.
ACCESSION AC090147
VERSION AC090147.3 GI:14114769
KEYWORDS HTG; HTGS_PHASE50.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

REFERENCE 1 (bases 1 to 6661)
Barren,B., Linton,L., Nusbaum,C. and Lander,E.
JOURNAL Unpublished
AUTHORS 2 (bases 1 to 6661)
Barren,B., Linton,L., Nusbaum,C., Lander,E., Allen,N., Anderson,S.,
Barna,N., Bastien,V., Boguslavsky,I., Bouckgalter,B., Brown,A.,
Camara,J., Campiano,A., Choepel,Y., Colangelo,M., Collins,S.,
Collamore,A., Cooke,P., Dearellano,K., Dewar,K., Diaz,J.S.,
Dodge,S., Fero,S., Ferreira,P., FitzHugh,W., Gege,D., Galagan,J.,
Gardyna,S., Ginde,S., Goyette,M., Graham,L., Grand-Pierre,N.,
Hagos,B., Hearford,A., Horton,L., Hulme,W., Iliev,I., Johnson,R.,
Jones,C., Karatas,A., LaRoque,K., Lamazares,R., Landers,T.,
Lehotzky,J., Levine,R., Liu,G., MacLean,C., Macdonald,P.,
Marquis,N., Matthews,C., McCarthy,M., McEwan,P., McKernan,K.,
McPheters,R., Meldrum,J., Menes,I., Minova,T.I., Mienga,V.,
Murphy,T., Naylor,J., Nguyen,C., Norbu,C., Norman,C.H.,
O'Connor,T., O'Donnell,P., O'Neill,D., Oliver,J., Peterson,K.,
Punkhang,P., Pierre.N., Pollara,V., Raymond,C., Retta,R.,
Rieback,M., Riley,R., Rise.C., Rogov.P., Roman.J., Rosetti,M.,
Roy,A., Santos.R., Schauer,S., Schnuppach,R., Seaman,S., Severy,P.,
Sugnez,C., Subramanian,A., Talamas,J., Testaye,S., Theodores,J.,
Travers.M., Travis,N., Triglio,J., Vassiliev,H., Viel,R., Vo,A.,
Wilson,B., Wu,X., Wyman.D., Ye.W.J., Young,G., Zainoun,J.,
Zembek,L., Zimmer.A. and Zody,M.

JOURNAL Submitted (17-FEB-2001) Whitehead Institute/MIT Center for Genome Direct Submission

TITLE JOURNAL

COMMENT

Research, 320 Charles Street, Cambridge, MA 02141, USA
 On May 18, 2001 this sequence version replaced gi:13324816.

All repeats were identified using RepeatMasker:

Smith, A.P.A. & Green, P. (1996-1997)
<http://ftp.genome.washington.edu/RM/RepeatMasker.html>

----- Genome Center

Center: Whitehead Institute/ MIT Center for Genome Research

Center code: WIBR

Web site: <http://www-seq.wi.mit.edu>

Contact: sequence_submissions@genome.wi.mit.edu

----- Project Information

Center project name: L11702

Center clone name: 382_A_18

* NOTE: This record contains 80 individual
 * sequencing reads that have not been assembled into
 * contigs. Runs of N are used to separate the reads
 * and the order in which they appear is completely
 * arbitrary. Low-pass sequence sampling is useful for
 * identifying clones that may be gene-rich and allows
 * overlap relationships among clones to be deduced.
 * However, it should not be assumed that this clone
 * will be sequenced to completion. In the event that
 * the record is updated, the accession number will
 * be preserved.

* 1 738: contig of 738 bp in length
 * 739 gap of 100 bp
 * 838: contig of 745 bp in length
 * 839 1583: contig of 745 bp in length
 * 1584 1683: gap of 100 bp
 * 1684 2428: contig of 745 bp in length
 * 2429 2528: gap of 100 bp
 * 2529 3247: contig of 719 bp in length
 * 3248 3347: gap of 100 bp
 * 3348 4079: contig of 732 bp in length
 * 4080 4179: gap of 100 bp
 * 4180 4888: contig of 709 bp in length
 * 4889 4988: gap of 100 bp
 * 4989 5717: contig of 729 bp in length
 * 5718 5817: gap of 100 bp
 * 5818 6571: contig of 754 bp in length
 * 6572 6671: gap of 100 bp
 * 6672 7424: contig of 753 bp in length
 * 7425 7524: gap of 100 bp
 * 7525 8260: contig of 736 bp in length
 * 8261 8360: gap of 100 bp
 * 8361 9094: contig of 734 bp in length
 * 9095 9194: gap of 100 bp
 * 9195 9914: contig of 720 bp in length
 * 9915 10014: gap of 100 bp
 * 10015 10739: contig of 725 bp in length
 * 10740 10839: gap of 100 bp
 * 10840 11567: contig of 728 bp in length
 * 11568 11667: gap of 100 bp
 * 11668 12390: contig of 723 bp in length
 * 12391 12490: gap of 100 bp
 * 12491 13218: contig of 728 bp in length
 * 13219 13318: gap of 100 bp
 * 13319 14021: contig of 703 bp in length
 * 14022 14121: gap of 100 bp
 * 14122 14846: contig of 725 bp in length
 * 14847 14946: gap of 100 bp
 * 14947 15683: contig of 737 bp in length
 * 15684 15783: gap of 100 bp
 * 15784 16526: contig of 743 bp in length
 * 16527 16626: gap of 100 bp
 * 16627 17357: contig of 731 bp in length
 * 17358 17457: gap of 100 bp
 * 17458 18196: contig of 739 bp in length
 * 18197 18296: gap of 100 bp
 * 18297 19015: contig of 719 bp in length
 * 19016 19115: gap of 100 bp
 * 19116 19856: contig of 741 bp in length
 * 19857 19956: gap of 100 bp

19957 20578: contig of 722 bp in length
 20679 20778: gap of 100 bp
 20779 21510: contig of 732 bp in length
 21510 21610: gap of 100 bp
 21611 22341: contig of 731 bp in length
 22341 22441: gap of 100 bp
 22442 23216: contig of 775 bp in length
 23216 23316: gap of 100 bp
 23317 24047: contig of 731 bp in length
 24048 24147: gap of 100 bp
 24148 24922: contig of 775 bp in length
 24923 25022: gap of 100 bp
 25023 25772: contig of 750 bp in length
 25773 25872: gap of 100 bp
 25873 26583: contig of 711 bp in length
 26584 27416: contig of 733 bp in length
 27417 27516: gap of 100 bp
 27517 28445: contig of 729 bp in length
 28446 28345: gap of 100 bp
 28346 29084: contig of 739 bp in length
 29085 29184: gap of 100 bp
 29185 29918: contig of 734 bp in length
 29919 30018: gap of 100 bp
 30019 30760: contig of 742 bp in length
 30761 30860: gap of 100 bp
 30861 31602: contig of 742 bp in length
 31603 31702: gap of 100 bp
 31703 32436: contig of 734 bp in length
 32437 32536: gap of 100 bp
 32537 33254: contig of 718 bp in length
 33255 33354: gap of 100 bp
 33355 34110: contig of 756 bp in length
 34111 34210: gap of 100 bp
 34211 34948: contig of 738 bp in length
 34949 35048: gap of 100 bp
 35049 35822: contig of 774 bp in length
 35823 35922: gap of 100 bp
 35923 36618: contig of 696 bp in length
 36619 36718: gap of 100 bp
 36719 37451: contig of 733 bp in length
 37452 37551: gap of 100 bp
 37552 38279: contig of 728 bp in length
 38280 38379: gap of 100 bp
 38380 39086: contig of 707 bp in length
 39087 39186: gap of 100 bp
 39187 39928: contig of 742 bp in length
 39929 40028: gap of 100 bp
 40029 40763: contig of 735 bp in length
 40764 40863: gap of 100 bp
 40864 41614: contig of 751 bp in length
 41615 41714: gap of 100 bp
 41715 42444: contig of 730 bp in length
 42445 42544: gap of 100 bp
 42545 43298: contig of 754 bp in length
 43299 43398: gap of 100 bp
 43399 44101: contig of 703 bp in length
 44102 44201: gap of 100 bp
 44202 44940: contig of 739 bp in length
 44941 45040: gap of 100 bp
 45041 45737: contig of 697 bp in length
 45738 45837: gap of 100 bp
 45838 46560: contig of 723 bp in length
 46561 47394: contig of 734 bp in length
 47395 47494: gap of 100 bp
 47495 48223: contig of 729 bp in length
 48224 48323: gap of 100 bp
 48324 49063: contig of 740 bp in length
 49064 49163: gap of 100 bp
 49164 49903: contig of 740 bp in length
 49904 50003: gap of 100 bp
 50004 50747: contig of 744 bp in length

* 50748 50847: gap of 100 bp
* 50848 51606: contig of 759 bp in length
* 51607 51706: gap of 100 bp
* 52462: contig of 756 bp in length
* 52463 52562: gap of 100 bp
* 52563 53294: contig of 732 bp in length
* 53295 53394: gap of 100 bp
* 53395 54138: contig of 744 bp in length
* 54139 54238: gap of 100 bp
* 54239 54956: contig of 718 bp in length
* 54957 55056: gap of 100 bp
* 55057 55791: contig of 735 bp in length
* 55792 55891: gap of 100 bp
* 55892 56624: contig of 733 bp in length
* 56625 56724: gap of 100 bp
* 56725 57456: contig of 732 bp in length

Query Match
Best Local Similarity 100.0%; Pred. No. 6.9e-65; Length 66611;
Matches 129; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1938 CCTAGACTAACGGTAATTTGGTGAAGAAAGAGTGAACACAGATTTTCTCAGTC 1997
|||||
Db 15977 CCTAGACTAACGGTAATTTGGTGAAGAAAGAGTGAACACAGATTTTCTCAGTC 15918
|||||

QY 1998 TGGAGAGAGCTTCAGATCCATCTCTGGAACAGTATTGATGTAGAGTTACCGAGC 2057
|||||
Db 15917 TGGAGAGAGCTTCAGATCCATCTCTGGAACAGTATTGATGTAGAGTTACCGAGC 15858
|||||

QY 2058 AGAATGGTA 2066
|||||
Db 15857 AGAATGGTA 15849
|||||

RESULT 10
AC090147 66611 bp DNA linear HTG 18-MAY-2001
LOCUS Homo sapiens chromosome 8 clone RP11-382A18 map 8, LOW-PASS
DEFINITION SEQUENCE SAMPLING.
AC090147
VERSION AC090147.3 GI:14141769
KEYWORDS HTG; HTGS PHASE0.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 66611)
AUTHORS Birren,B., Linton,L., Nusbaum,C. and Lander,E.
TITLE Homo sapiens chromosome 8, clone RP11-382A18
JOURNAL Unpublished
AUTHORS
REFERENCE 2 (bases 1 to 66611)
AUTHORS Birren,B., Linton,L., Nusbaum,C., Lander,E., Allen,N., Anderson,S.,
Barna,N., Bastien,V., Boguslavsky,L., Boukhgalter,B., Brown,A.,
Camata,J., Campopiano,A., Choepel,Y., Colangelo,M., Collins,S.,
Collamore,A., Cooke,P., Dearellano,K., Dewar,K., Diaz,J.S.,
Dodge,S., Fairo,S., Ferreira,P., FitzHugh,W., Gage,D., Galagan,J.,
Gardyna,S., Ginde,S., Goyette,M., Graham,L., Grand-Pierre,N.,
Hagos,B., Heaford,A., Horton,L., Hulme,W., Iliev,I., Johnson,R.,
Jones,C., Karatas,A., LaRocque,K., Lamazares,R., Landers,T.,
Lehoczky,J., Levine,R., Liu,G., MacLean,C., Macdonald,P.,
Marquis,N., Matthews,C., McCarthy,M., McEwan,P., McKernan,K.,
McPheters,R., Meldrim,J., Meneus,L., Mihova,T., Mlenga,V.,
Murphy,T., Naylor,J., Nguyen,C., Norbu,C., Norman,C.H.,
O'Connor,T., O'Donnell,P., O'Neill,D., Oliver,J., Peterson,K.,
Phunkhang,P., Pierre,N., Pollara,V., Raymond,C., Retta,R.,
Rieckback,M., Riley,R., Rise,C., Rogov,P., Roman,J., Rosetti,M.,
Roy,A., Santos,R., Schauer,S., Schuback,R., Seaman,S., Severy,P.,
Sougnuez,C., Spencer,B., Stange-Thomann,N., Stojanovic,N.,
Strauss,N., Subramanian,A., Talamas,J., Tesfaye,S., Theodore,J.,
Travers,M., Travis,N., Trigilio,J., Vassiliev,H., Viel,R., Vo,A.,
Wilson,B., Wu,X., Wyman,D., Ye,W.J., Young,G., Zainoun,J.,
Zembek,L., Zimmer,A. and Zody,M.
Direct Submission

JOURNAL Submitted (17-FEB-2001) Whitehead Institute/MIT Center for Genome
Research, 320 Charles Street, Cambridge, MA 02141, USA
COMMENT On May 18, 2001 this sequence version replaced gi:13324816.
All repeats were identified using RepeatMasker:
Smit, A.F.A. & Green, P. (1996-1997)
http://ftp.genome.washington.edu/RM/RepeatMasker.html
----- Genome Center
Center: Whitehead Institute/ MIT Center for Genome Research
Center code: WIER
Web site: http://www-seq.wi.mit.edu
Contact: sequence_submissions@genome.wi.mit.edu
----- Project Information
Center project name: L11702
Center clone name: 382_A_18

* NOTE: This record contains 80 individual
* sequencing reads that have not been assembled into
* contigs. Runs of N are used to separate the reads
* and the order in which they appear is completely
* arbitrary. Low-pass sequence sampling is useful for
* identifying clones that may be gene-rich and allows
* overlap relationships among clones to be deduced.
* However, it should not be assumed that this clone
* will be sequenced to completion. In the event that
* the record is updated, the accession number will
* be preserved.

* 1 738: contig of 738 bp in length
* 739 838: gap of 100 bp
* 839 1583: contig of 745 bp in length
* 1584 1683: gap of 100 bp
* 1684 2428: contig of 745 bp in length
* 2429 2528: gap of 100 bp
* 2529 3247: contig of 719 bp in length
* 3248 3348: gap of 100 bp
* 3349 4079: contig of 732 bp in length
* 4080 4179: gap of 100 bp
* 4180 4888: contig of 709 bp in length
* 4889 4988: gap of 100 bp
* 4989 5717: contig of 729 bp in length
* 5718 5817: gap of 100 bp
* 5818 6571: contig of 754 bp in length
* 6572 6671: gap of 100 bp
* 6672 7424: contig of 753 bp in length
* 7425 7524: gap of 100 bp
* 7525 8260: contig of 736 bp in length
* 8261 8360: gap of 100 bp
* 8361 9094: contig of 734 bp in length
* 9095 9194: gap of 100 bp
* 9195 9914: contig of 720 bp in length
* 9915 10014: gap of 100 bp
* 10015 10739: contig of 725 bp in length
* 10740 10839: gap of 100 bp
* 10840 11567: contig of 728 bp in length
* 11568 12390: contig of 723 bp in length
* 12391 12490: gap of 100 bp
* 12491 13218: contig of 728 bp in length
* 13219 13318: gap of 100 bp
* 13319 14021: contig of 703 bp in length
* 14022 14121: gap of 100 bp
* 14122 14846: contig of 725 bp in length
* 14847 14946: gap of 100 bp
* 14947 15683: contig of 737 bp in length
* 15684 15783: gap of 100 bp
* 15784 16526: contig of 743 bp in length
* 16527 16626: gap of 100 bp
* 16627 17357: contig of 731 bp in length
* 17358 17457: gap of 100 bp
* 17458 18136: contig of 739 bp in length
* 18137 18296: gap of 100 bp
* 18297 19015: contig of 719 bp in length
* 19016 19115: gap of 100 bp
* 19116 19856: contig of 741 bp in length

```
* 19557 19556: gap of 100 bp
* 19557 20678: contig of 722 bp in length
* 20678 20778: gap of 100 bp
* 20778 21510: contig of 732 bp in length
* 21510 21610: gap of 100 bp
* 21610 22341: contig of 731 bp in length
* 22341 22441: gap of 100 bp
* 22441 23216: contig of 775 bp in length
* 23216 23316: gap of 100 bp
* 23316 24047: contig of 731 bp in length
* 24047 24147: gap of 100 bp
* 24147 24922: contig of 775 bp in length
* 24922 25022: gap of 100 bp
* 25022 25772: contig of 750 bp in length
* 25772 25872: gap of 100 bp
* 25872 26583: contig of 711 bp in length
* 26583 26683: gap of 100 bp
* 26683 27416: contig of 733 bp in length
* 27416 27516: gap of 100 bp
* 27516 28245: contig of 729 bp in length
* 28245 28345: gap of 100 bp
* 28345 29084: contig of 739 bp in length
* 29084 29184: gap of 100 bp
* 29184 29918: contig of 734 bp in length
* 29918 30018: gap of 100 bp
* 30018 30760: contig of 742 bp in length
* 30760 30860: gap of 100 bp
* 30860 31602: contig of 742 bp in length
* 31602 31702: gap of 100 bp
* 31702 32436: contig of 734 bp in length
* 32436 32536: gap of 100 bp
* 32536 33254: contig of 718 bp in length
* 33254 33354: gap of 100 bp
* 33354 34110: contig of 756 bp in length
* 34110 34210: gap of 100 bp
* 34210 34948: contig of 738 bp in length
* 34948 35048: gap of 100 bp
* 35048 35822: contig of 774 bp in length
* 35822 35922: gap of 100 bp
* 35922 36618: contig of 696 bp in length
* 36618 36718: gap of 100 bp
* 36718 37451: contig of 733 bp in length
* 37451 37551: gap of 100 bp
* 37551 38279: contig of 728 bp in length
* 38279 38379: gap of 100 bp
* 38379 39086: contig of 707 bp in length
* 39086 39186: gap of 100 bp
* 39186 39288: contig of 742 bp in length
* 39288 40028: gap of 100 bp
* 40028 40763: contig of 735 bp in length
* 40763 40863: gap of 100 bp
* 40863 41614: contig of 751 bp in length
* 41614 41714: gap of 100 bp
* 41714 42444: contig of 730 bp in length
* 42444 42544: gap of 100 bp
* 42544 43298: contig of 754 bp in length
* 43298 43398: gap of 100 bp
* 43398 44101: contig of 703 bp in length
* 44101 44201: gap of 100 bp
* 44201 44940: contig of 739 bp in length
* 44940 45040: gap of 100 bp
* 45040 45737: contig of 697 bp in length
* 45737 45837: gap of 100 bp
* 45837 46560: contig of 723 bp in length
* 46560 46660: gap of 100 bp
* 46660 47394: contig of 734 bp in length
* 47394 47494: gap of 100 bp
* 47494 48223: contig of 729 bp in length
* 48223 48323: gap of 100 bp
* 48323 49063: contig of 740 bp in length
* 49063 49163: gap of 100 bp
* 49163 49904 49903: contig of 740 bp in length
* 49904 50003: gap of 100 bp

* 50004 50747: contig of 744 bp in length
* 50747 50847: gap of 100 bp
* 50847 51606: contig of 759 bp in length
* 51606 51706: gap of 100 bp
* 51706 52462: contig of 756 bp in length
* 52462 52562: gap of 100 bp
* 52562 53294: contig of 732 bp in length
* 53294 53394: gap of 100 bp
* 53394 54138: contig of 744 bp in length
* 54138 54238: gap of 100 bp
* 54238 54956: contig of 718 bp in length
* 54956 55056: gap of 100 bp
* 55056 55791: contig of 735 bp in length
* 55791 55891: gap of 100 bp
* 55891 56624: contig of 733 bp in length
* 56624 56724: gap of 100 bp
* 56724 57456: contig of 732 bp in length
* 57456 57525 57456: contig of 732 bp in length

Query Match 5.4%; Score 125; DB.2; Length 66611;
Best Local Similarity 100.0%; Fred. No. 1.8e-62;
Matches 125; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 272 AGGTTACTGAAGTTTGTAGTAAGACCCGGAATGAGAGGATGTGAGGATACGATCACTT 331
Db 45398 AGGTTACTGAAGTTTGTAGTAAGACCCGGAATGAGAGGATGTGAGGATACGATCACTT 45457

QY 332 TAACAAATGAATTCACCTACATACCACTTGTGAGTCTATATATATTTTCA 391
Db 45458 TAACAAATGAATTCACCTACATACCACTTGTGAGTCTATATATATTTTCA 45517

QY 392 GGAGG 396
Db 45518 GGAGG 45522

RESULT 11
BD114143
LOCUS BD114143 461 bp DNA linear PAT 18-SEP-2002
DEFINITION EST and encoded human protein.
ACCESSION BD114143
VERSION BD114143.1 GI:23209047
KEYWORDS JP 2002010789-A/6220.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 461)
AUTHORS Edwards,J.B.D.M., Jobert,S. and Giordano,J.E.
TITLE EST and encoded human protein
JOURNAL Patent: JP 2002010789-A 6220 15-JAN-2002;
COMMENT GENSER CORP
OS Homo sapiens (human)
PN JP 2002010789-A/6220
PD 15-JAN-2002
PF 07-AUG-2000 JP 2000280989
PR 05-AUG-1999 US 60/147499
PI JEAN EAPUTIST DUMAS MILNE EDWARDS, SEVELIN JOBERT, JEAN EVE PI
GIORDANO
PC C12N15/09,C12N15/09,C07K14/47,C07K16/18,C12N1/15,C12N1/19, PC
C12N1/21,
PC C12N5/10,C12P21/02,C12P21/08,C12Q1/69,C12N15/00,C12N5/00, PC
C12N15/00
CC n=a, g, c or t Location/Qualifiers
FH Key Location/Qualifiers
FT misc feature 107.
FEATURES
source
1..461
/organism="Homo sapiens"
/mol_type="genomic DNA"
/db_xref="taxon:9606"
BASE COUNT 136 a 107 c 129 g 87 t 2 others
ORIGIN
```

```
Query Match      4.0%; Score 92; DB 6; Length 461;
Best Local Similarity 100.0%; Pred. No. 1.2e-42;
Matches 92; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 32 TAGACCATTGTAAGAAATCAAAACAGGTTCTTCAGGCATTATAGTAAGGTTAAGCACTA 91
   |||||
Db 335 TAGACCATTGTAAGAAATCAAAACAGGTTCTTCAGGCATTATAGTAAGGTTAAGCACTA 394
   |||||

QY 92 ACATTTCGGCTGACATCCCGTCCCGAGTGG 123
   |||||
Db 395 ACCATTTCGGCTGACATCCCGTCCCGAGTGG 426
   |||||

RESULT 12
AX209866/c
LOCUS      460 bp      DNA      linear      PAT 31-AUG-2001
DEFINITION Sequence 1706 from Patent WO0157207.
ACCESSION  AX209866
VERSION     AX209866.1 GI:15424289
KEYWORDS
SOURCE      Homo sapiens (human)
ORGANISM    Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1
AUTHORS    Algate, P. A. and Mannion, J.
TITLE      Compositions and methods for the therapy and diagnosis of ovarian
JOURNAL    Cancer
PATENT     WO 0157207-A 1706 09-AUG-2001;
CORIXA     CORPORATION (US)
FEATURES    Location/Qualifiers
            1..460
            /organism="Homo sapiens"
            /mol_type="genomic DNA"
            /db_xref="taxon:9606"
BASE COUNT  188 a 78 c 69 g 125 t
ORIGIN

Query Match      2.5%; Score 58; DB 6; Length 460;
Best Local Similarity 100.0%; Pred. No. 3.9e-22;
Matches 58; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2271 CCAGAGTATTTCAGAGAGCCAAATCTGTCACTGTCTCAACGGCTTTACTACCTCTAA 2328
   |||||
Db 460 CCAGAGTATTTCAGAGAGCCAAATCTGTCACTGTCTCAACGGCTTTACTACCTCTAA 403
   |||||

RESULT 13
AB032604
LOCUS      3910 bp      mRNA      linear      ROD 04-OCT-2001
DEFINITION Mus musculus mRNA for MIWI, complete cds.
ACCESSION  AB032604
VERSION     AB032604.1 GI:7416110
KEYWORDS    MIWI (piwi).
SOURCE      Mus musculus (house mouse)
ORGANISM    Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1
AUTHORS    Kuramochi-Miyagawa, S., Kimura, T., Yomogida, K., Kuroiwa, A.,
            Tadokoro, Y., Fujita, Y., Sato, M., Matsuda, Y. and Nakano, T.
TITLE      Two mouse piwi-related genes: miwi and mili
JOURNAL    Mech Dev. 108 (1-2), 121-133 (2001)
MEDLINE    21463379
PUBMED     11578866
REFERENCE   2 (bases 1 to 3910)
AUTHORS    Miyagawa, S.K. and Nakano, T.
TITLE      Direct Submission
JOURNAL    Submitted (21-SEP-1999) Satomi K Miyagawa, Research Institute for
            Microbial Diseases, Department of Molecular Cell Biology; Yamadaoka
            3-1, Suita, Osaka 565-0871, Japan
            (E-mail: smiyagawa@biken.osaka-u.ac.jp, Tel:81-6-6879-8362,
            Fax:81-6-6879-8362)

Query Match      1.5%; Score 35; DB 10; Length 3910;
Best Local Similarity 100.0%; Pred. No. 3.1e-08;
Matches 35; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1687 GCACGATCAATGAGGGATGACCCGCTGGTCTC 1721
   |||||
Db 2103 GCACGATCAATGAGGGATGACCCGCTGGTCTC 2137
   |||||

RESULT 14
AF438405
LOCUS      4064 bp      mRNA      linear      ROD 13-NOV-2001
DEFINITION Mus musculus MIWI mRNA, complete cds.
ACCESSION  AF438405
VERSION     AF438405.1 GI:16905060
KEYWORDS
SOURCE      Mus musculus (house mouse)
ORGANISM    Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1 (bases 1 to 4064)
AUTHORS    Deng, W. and Lin, H.
TITLE      Miwi, a murine homolog of piwi, encodes a cytoplasmic protein
JOURNAL    Unpublished
REFERENCE   2 (bases 1 to 4064)
AUTHORS    Deng, W. and Lin, H.
TITLE      Direct Submission
JOURNAL    Submitted (22-OCT-2001) Department of Cell Biology, Duke
            University, Room 412, NanaLine Duke Building, Durham, NC 27710, USA
FEATURES    Location/Qualifiers
            1..4064
            /organism="Mus musculus"
            /mol_type="mRNA"
            /db_xref="taxon:10090"
            /chromosome="5"
            /tissue_type="testis"
            191..2779
BASE COUNT  993 a 970 c 1063 g 884 t
ORIGIN
```

```
Location/Qualifiers
1..3910
/organism="Mus musculus"
/mol_type="mRNA"
/db_xref="taxon:10090"
/map="5p distal-G2 proximal"
/tissue_lib="testis"
1..3910
/gene="miwi"
156..2744
/gene="miwi"
/note="piwi family"
/codon_start=1
/product="MIWI"
/protein_id="BAA93705.1"
/db_xref="GI:7416111"
/translation="MTGRARARARARAGQETVQHVHGAASQPCGYIPRPOOSPTEG
DLVGRGRQGVVGVGATSKSQELQISAGFQELSLAERGRRRDFHLGVNTRQNLQDLVGR
ESKTSSGIIIVKLSTNHLTSRQWALYQHDYDYNLEARRKRSALLFQHEDLIGR
CHAFDGTLLFLPKLQHKVTEVFSQTRNGEVRITITLNLPPPTCLOFYNIIFR
RLKIMNLOQIGRNYNEPDPIDIPNHLVIVPGFTTSLQYENNMILCTDYSKVLK
SETVLDEWNLVQOTERHFKFORVSKELIGLIVTKYNNKTVYRDDIDWDQPKSTFK
KADGSEVSEFLVYKQYNOEITDLKQPVLSQPKERRGGGTLPGPAMLIPELCVLTG
LTDKRNDFNVNKLAVHRLTPEQOREVSEGLIDYIHKDDNVQRELDWGLSFDNSL
LSFSGRILOSEKIHQGGKTFDYPQFADWSKETRGAPLSVPLDNWLLIYTRNEYA
ANSLIQNLKVTTPAMGIQMKAIMTEVDRTAEYLRALQCKVTSITQIVVCLSSNRK
DKYDAIKKYLCTDCPTSPQCVVARTLQKQVTMAIATKALQNMKGELWRVDMPL
KLAMTVGIDCYHDTAGRSIVAGFVASINEGMTWFSRCVQDRQOGLVDGLKVLQIA
ALRAMSGCNEVHPSRVIVRQGVGQGLKLVNVEVPOFLDCLKVGVRGRLTVIV
VKRVNARFFAQSGLRQNLPLEGTVIDVEVPEWYDFIVSQAVRSGSVSTHYNVI
YDSSGLKPDHIOQLTYIKLCHVYNNPVGIVRPAQYAHKLAFLVQGSIHREPNSLS
NRLYYL"
BASE COUNT  993 a 970 c 1063 g 884 t
ORIGIN

Query Match      1.5%; Score 35; DB 10; Length 3910;
Best Local Similarity 100.0%; Pred. No. 3.1e-08;
Matches 35; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1687 GCACGATCAATGAGGGATGACCCGCTGGTCTC 1721
   |||||
Db 2103 GCACGATCAATGAGGGATGACCCGCTGGTCTC 2137
   |||||

RESULT 14
AF438405
LOCUS      4064 bp      mRNA      linear      ROD 13-NOV-2001
DEFINITION Mus musculus MIWI mRNA, complete cds.
ACCESSION  AF438405
VERSION     AF438405.1 GI:16905060
KEYWORDS
SOURCE      Mus musculus (house mouse)
ORGANISM    Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1 (bases 1 to 4064)
AUTHORS    Deng, W. and Lin, H.
TITLE      Miwi, a murine homolog of piwi, encodes a cytoplasmic protein
JOURNAL    Unpublished
REFERENCE   2 (bases 1 to 4064)
AUTHORS    Deng, W. and Lin, H.
TITLE      Direct Submission
JOURNAL    Submitted (22-OCT-2001) Department of Cell Biology, Duke
            University, Room 412, NanaLine Duke Building, Durham, NC 27710, USA
FEATURES    Location/Qualifiers
            1..4064
            /organism="Mus musculus"
            /mol_type="mRNA"
            /db_xref="taxon:10090"
            /chromosome="5"
            /tissue_type="testis"
            191..2779
BASE COUNT  993 a 970 c 1063 g 884 t
ORIGIN
```

```
/note="Piwi-like protein"
/codon_start=1
/product="MIWI"
/protein_id="AAL31014.1"
/db_xref="GI:16905061"
/translation="MTGRARARARGAQETVHVGAASQPGYIPRPOOSPTEG
DLVGRQKQMVVGTAKSQELQISAGFQELSLAERGRRRPHDLGVNTRQNLDAVK
ESKTSGGIIVKLSTNHFRLTSPQWALYQYHIDYNPLMEARLRALLFQHEDLIGR
CHAFDTIILFLPRQLQKVTVEFSQTRNGEHRVITITLNELPSTPTCLOFYNIIFR
RLIKMNLQOIGNYNPSDPIIDPNHRLVINGPFTSILOYENIMLCTDVSHKVLIR
SEVLDPFENLYQTEEHKFOEQVSKELIGLIVLTKNKTYRVDIDWDONPKSTFK
KADGSEVSEFLYRKQYNOBELTLKQPLVSPQKRRGPGGTLPGEMLI BELCVLTG
LTDKMRDNFMDLAVHRLTLEPQREVRGLDIYHKDDNVQRELRDWGLSFDNL
LSFSGRILOSEKIHQGGKTFDYNPQFADWSKETRGAPLISVKELDNWLLIYTRNTEA
ANSILQNLKFTVTAMGIQMKKAIMIEVDRTTEAYLRALQOKVTSDTQIVVCLSSNRK
KYDAIKYLCCTPPTSCQWARTIGKQOTVMAIATKIALQWNCMGGLMRVDMPL
KLAMIVIGDYCHYTTAGRRSIAGFVASINEGTRFRCVFDQGOELVDGLKVCLOA
ALAWSGCNEYMSRVLVYEDGQGLKTLVNYEVPQELDLCKSVGRGVNPELTVIV
VKXRVNARPPAQGGGLQNPGLPGTVIDVEVTRPEWIDFFIVSOAVRSGVSPHYNVI
YDSGLKPDHIOIRLTLYKLCHVYNNWPGVIRVPAPQYAHKLAFLVQGSIHREPNLSLS
NRLYYL"
```

```
BASE COUNT 1115 a 982 c 1077 g 890 t
ORIGIN
Query Match 1.5%; Score 35; DB 10; Length 4064;
Best Local Similarity 100.0%; Pred. No. 3.1e-08;
Matches 35; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

```
QY 1687 GCCAGCATCAATGAAGGATGACCCGCTGGTTCTC 1721
Db 2138 GCCAGCATCAATGAAGGATGACCCGCTGGTTCTC 2172

RESULT 15
AX395819 501 bp DNA linear PAT 18-MAY-2002
LOCUS AX395819
DEFINITION Sequence 34 from Patent WO0212328.
ACCESSION AX395819
VERSION AX395819.1 GI:21066566
KEYWORDS
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1
AUTHORS King, G.E., Meagher, M.J., Xu, J. and Secrist, H.
TITLE Compositions and methods for the therapy and diagnosis of colon
cancer
JOURNAL Parent: WO 0212328-A 34 14-FEB-2002;
CORIXA CORPORATION (US)
FEATURES
source
1..501
/organism="Homo sapiens"
/mol_type="genomic DNA"
/db_xref="taxon:9606"
BASE COUNT 134 a 81 c 93 g 193 t
ORIGIN
```

```
Query Match 1.5%; Score 34; DB 6; Length 501;
Best Local Similarity 100.0%; Pred. No. 1.2e-07;
Matches 34; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

```
QY 2271 CCAGAGTATTCACAGAGAGCCAAATCTCTCACTG 2304
Db 1 CCAGAGTATTCACAGAGAGCCAAATCTCTCACTG 34
```

Search completed: December 6, 2003, 20:44:23
Job time : 8515 secs

Result No.	Score	Query Match	Length	DB ID	Description
1	2042	87.7	3472	21 AAA07588	Human piwi gene, d
2	1569	67.4	2292	25 ABV99287	Human colon cancer
3	316	13.6	367	25 ABZ19511	Group III cDNA can
4	284	12.2	300	20 AAZ13404	Human gene express
5	269	11.6	423	21 AAF21880	Human breast and o
6	214	9.2	367	25 ABZ20005	Group III cDNA can
7	158	6.8	209	25 ABZ19102	Group III cDNA can
8	60	2.6	60	24 ABN42800	Human spliced tran

1832	TCATCGTGTAACCGCGATGGCGTAGAGGAGACGGCCAGCTGAGAAAACACTGGTGAACTACGAAG	1891
2256	TCATCGTGTAACCGCGATGGCGTAGGAGAGCGCCAGCTGAAAACTCTGGTGAACTACGAAG	2315
1892	TGCCACAGTTTTTTGGATTTGTTCTAAAATCCATTTGGTAGAGTTACAACCCCTAGACTAACGG	1951
2316	TGCCACAGTTTTTTGGATTTGTTCTAAAATCCATTTGGTAGAGTTACAACCCCTAGANTAACGG	2375
1952	TAAATTGCTGAGAAAAGAGTCAACACCCAGATTTTTTGTCTCAGTCTGGAGGGAAGACTTC	2011
2376	TAAATTGCTGAGAAAAGAGTCAACACCCAGATTTTTTGTCTCAGTCTGGAGGGAAGACTTC	2435
2012	AGAAATCCACTTCCTGGAAACAGTTATTGATGTAGAGGTTACACAGACCCAGAAATGCTATGACT	2071
2436	AGAAATCCACTTCCTGGAAACAGTTATTGATGTAGAGGTTACACAGACCCAGAAATGCTATGACT	2495
2072	TTTTTTATCGTGAGCCAGGCTGTGAGAAAGTGGTAGTGTTCCTCCACACATTACAATGTCA	2131
2496	TTTTTTATCGTGAGCCAGGCTGTGAGAAAGTGGTAGTGTTCCTCCACACATTACAATGTCA	2555
2132	TCTATGACAAACAGGGGCTGAAGCCACACACATACAGCGCTTGACCTACAAGCTGTGCC	2191
2556	TCTATGACAAACAGGGGCTGAAGCCACACACATACAGCGCTTGACCTACAAGCTGTGCC	2615
2192	ACATCTATTATCAACTGGCCAGGTGTCAATCGTGTTCCTGTCTCCTTGCCAGCTAGCCCCACA	2251
2616	ACATCTATTATCAACTGGCCAGGTGTCAATCGTGTTCCTGTCTCCTTGCCAGCTAGCCCCACA	2675
2252	AGCTGGCTTTTCTTGTGTGGCCAGAGTATTCAACAGAGAGCCAAATCTGTCTACTGTCAAACC	2311
2676	AGCTGGCTTTTCTTGTGTGGCCAGAGTATTCAACAGAGAGCCAAATCTGTCTACTGTCAAACC	2735
2312	GCCTTTTACTACCTCTAA	2328
2736	GCCTTTTACTACCTCTAA	2752

RESULT 2

ABV89287
 ID ABV89287 standard; cDNA; 2292 BP.
 XX
 XX ABV89287;
 AC
 XX
 DT 13-DEC-2002 (first entry).
 XX
 XX Human colon cancer related cDNA SEQ ID NO 2602.
 DE
 XX
 XX Human; colon; cancer; cytostatic; tumour; gene therapy; vaccine; gene;
 KW ss.
 XX
 XX Homo sapiens.
 OS
 XX
 PN WO200258534-A2.
 XX
 XX 01-AUG-2002.
 ED
 XX
 XX 19-NOV-2001; 2001WO-US43704.
 PF
 XX
 XX 20-NOV-2000; 2000US-252222P.
 PR 06-FEB-2001; 2001US-267011P.
 PR 28-MAR-2001; 2001US-279670P.
 PR 10-JUL-2001; 2001US-304037P.
 PR
 PA (CORI-) CORIXA CORP.
 XX
 XX Stolk JA, Xu J, Chenault RA, Meagher MJ, Secrist H, King GE;
 PI
 XX
 XX WPI: 2002-608400/65.
 DR P-PSDE; ABP67395.
 DR
 XX
 XX
 PT New isolated tumor colon polynucleotide and polypeptide, useful for the
 PT diagnosis, prevention and/or treatment of cancer, in particular colon
 PT cancer -

XX	PS	Claim 1; SEQ ID NO 2602; 266pp + Sequence Listing; English.
XX	CC	The invention relates to a human colon tumour expressed polynucleotide
CC	CC	(I) encoding a polypeptide (II, ABP67991-ABP67996) comprising: (i) any of
CC	CC	2600 fully defined nucleotide sequences (ABV8669-ABV89289); (ii)
CC	CC	complements of (i); (iii) at least 20 contiguous residues of (i); (iv)
CC	CC	sequences that hybridize to (i), under moderately stringent conditions;
CC	CC	(v) sequences having at least 75% or 90% identity to (i); or (vi)
CC	CC	degenerate variants of (i). The compositions and methods of the present
CC	CC	invention are useful for the diagnosis, prevention and/or treatment of
CC	CC	cancer, particularly colon cancer. (I) can be used in gene therapy and
CC	CC	(I) and (II) are useful in pharmaceutical compositions such as vaccines.
CC	CC	Note: The sequence data for this patent did not form part of the printed
CC	CC	specification, but was obtained in electronic format directly from WIPO
CC	CC	at ftp.wipo.int/pub/published_pct_sequences .
XX	XX	
XX	XX	Sequence 2292 BP; 692 A; 435 C; 512 G; 653 T; 0 other;
XX	XX	
XX	XX	Query Match 67.4%; Score 1569; DB 24; Length 2292;
XX	XX	Best Local Similarity 100.0%; Pred. No. 0;
XX	XX	Matches 1569; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
XX	XX	
QY	Db	760 GTGAGCTTCTTAGAATACTACGAGAGCAATACACCAAGAGATCACCGACTTTGAAGCAG 819
QY	Db	4 GTGAGCTTCTTAGAATACTACGAGAGCAATACACCAAGAGATCACCGACTTTGAAGCAG 63
QY	Db	820 CTTGTCTTTGTGTCAGCCAGCCCAAGAGAGCGGGGCCCTGGGGGACACTGCGCAGGGCCT 879
QY	Db	64 CTTGTCTTTGTGTCAGCCAGCCCAAGAGAGCGGGGCCCTGGGGGACACTGCGCAGGGCCT 123
QY	Db	880 GCCATGCTCANTCCTGAGCTCTGCTATCTTACAGGCTTAACTGATAAATGCGTAATGAT 939
QY	Db	124 GCCATGCTCANTCCTGAGCTCTGCTATCTTACAGGCTTAACTGATAAATGCGTAATGAT 183
QY	Db	940 TTTAACTGTATGAAGACTTTAGCCGTTTCATCAAGACTTAACTCCAGAGCAAAAGCAGCGT 999
QY	Db	184 TTTAACTGTATGAAGACTTTAGCCGTTTCATCAAGACTTAACTCCAGAGCAAAAGCAGCGT 243
QY	Db	1000 GAAGTGGGAGCGACTCAATTGATTATCAATAAAGCATATGTTCAAAGGAGCCTTCGA 1059
QY	Db	244 GAAGTGGGAGCGACTCAATTGATTATCAATAAAGCATATGTTCAAAGGAGCCTTCGA 303
QY	Db	1060 GACTGGGGTTTGAGCTTTGATTCGAACTTACTGTCTTCTCAGAGAGAAATTTTGCAGAAC 1119
QY	Db	304 GACTGGGGTTTGAGCTTTGATTCGAACTTACTGTCTTCTCAGAGAGAAATTTTGCAGAAC 363
QY	Db	1120 GAAAGATTACCAAGGTGGGAAAACATTTGATTACAATCCACAATTTGCAGATTTGGTCC 1179
QY	Db	364 GAAAGATTACCAAGGTGGGAAAACATTTGATTACAATCCACAATTTGCAGATTTGGTCC 423
QY	Db	1180 AAAGAAAACAAGGATGCGACCAATTAAATAGTGTTAAGCCACTAGATAAATCTGCTGTGATC 1239
QY	Db	424 AAAGAAAACAAGGATGCGACCAATTAAATAGTGTTAAGCCACTAGATAAATCTGCTGTGATC 483
QY	Db	1240 TATACGGAAGAAATATGAGCGAGCCAAATTCATTGATACAAAATCTATTTAAAGTTACA 1299
QY	Db	484 TATACGGAAGAAATATGAGCGAGCCAAATTCATTGATACAAAATCTATTTAAAGTTACA 543
QY	Db	1300 CCAGCCATGGCCATGCAATCAGAAAAAGCAATAATGATTGAAGTGGATGACAGAACTCAA 1355
QY	Db	544 CCAGCCATGGCCATGCAATCAGAAAAAGCAATAATGATTGAAGTGGATGACAGAACTCAA 603
QY	Db	1360 GCCTACTTAAAGAGTCTTACGCAAAAAGGTTCACGACAGACCCAGATAGTTGTCTGTCTG 1419
QY	Db	604 GCCTACTTAAAGAGTCTTACGCAAAAAGGTTCACGACAGACCCAGATAGTTGTCTGTCTG 663
QY	Db	1420 TTGTCAAGTAATCGGAAGGACAAATACGATCTATTTAAAAATACCTGTGTACAGATTGC 1479
QY	Db	664 TTGTCAAGTAATCGGAAGGACAAATACGATCTATTTAAAAATACCTGTGTACAGATTGC 723
QY	Db	1480 CCTACCCCAAGTCAGTGTGTGTGGCCCGAACCTTAGCGCAACAGCAACTGTCAATGCC 1539

Db 724 CTTACCCCAAGTCAGTGTGTGTGTGGCCGGAACCTTAGGCAAAACAGCAAACTGTCTGGCC 783
Qy 1540 ATTGCTTACAAAGATTCCCTTACAGATGAATCTGCAAGATGGGAGAGAGCTCTGGAGGGTG 1599
Db 784 ATTGCTTACAAAGATTCCCTTACAGATGAATCTGCAAGATGGGAGAGAGCTCTGGAGGGTG 843
Qy 1600 GACATCCCTCTGAAGCTGTGTATGATGCTGTGGCATCGATTGTACCATGATGACAGCT 1659
Db 844 GACATCCCTCTGAAGCTGTGTATGATGCTGTGGCATCGATTGTACCATGATGACAGCT 903
Qy 1660 GGGCGAGGTCAATCGCAGATTTGTTGCCAGCATCAATGAAGGGATGACCCCTGGTTC 1719
Db 904 GGGCGAGGTCAATCGCAGATTTGTTGCCAGCATCAATGAAGGGATGACCCCTGGTTC 963
Qy 1720 TCACGCTGCATATTTTACAGATAGAGGACAGAGCTGTGTAGATGGCTCAAGTCTGCCTG 1779
Db 964 TCACGCTGCATATTTTACAGATAGAGGACAGAGCTGTGTAGATGGCTCAAGTCTGCCTG 1023
Qy 1780 CAAGCGCTCTGAGGGCTTGGAAATAGCTGCAATGAGTACATGCCCGAGATCATCGTG 1839
Db 1024 CAAGCGCTCTGAGGGCTTGGAAATAGCTGCAATGAGTACATGCCCGAGATCATCGTG 1083
Qy 1840 TACCGGATCGGTAGAGAGCGCCAGCTGAAACACTGTGTGACTGACTGACGAGTGCACAG 1899
Db 1084 TACCGGATCGGTAGAGAGCGCCAGCTGAAACACTGTGTGACTGACTGACGAGTGCACAG 1143
Qy 1900 TTTTGTGATTGTCTTAAATCATTTGAGAGTTTACCACTTACCACTGAGTAAATGTTG 1959
Db 1144 TTTTGTGATTGTCTTAAATCATTTGAGAGTTTACCACTTACCACTGAGTAAATGTTG 1203
Qy 1960 GTGAAGAAAAGAGTGAACACAGATTTTGTCTCAGTCTGAGGAGACACTTCAAGATCCA 2019
Db 1204 GTGAAGAAAAGAGTGAACACAGATTTTGTCTCAGTCTGAGGAGACACTTCAAGATCCA 1263
Qy 2020 CTTCTGGAACAGTATTGATGATGAGTTTACGAGCAGCAATGATGATGATGATGATGATG 2079
Db 1264 CTTCTGGAACAGTATTGATGATGAGTTTACGAGCAGCAATGATGATGATGATGATGATG 1323
Qy 2080 GTGAGCAGGCTGTGAGAGTGTGTGTCTTCTCCACACATTAACAATGTCATCTATGAC 2139
Db 1324 GTGAGCAGGCTGTGAGAGTGTGTGTCTTCTCCACACATTAACAATGTCATCTATGAC 1383
Qy 2140 AACAGCGCTGTGAAGCCAGACCAATACAGCGTGTGACCTTACAGCTGTGACCACTTAT 2199
Db 1384 AACAGCGCTGTGAAGCCAGACCAATACAGCGTGTGACCTTACAGCTGTGACCACTTAT 1443
Qy 2200 TACACTGGCCAGGTGTCTTGTGTTCTCTCTCTCTGAGTACGCGCCCAAGCTGGCT 2259
Db 1444 TACACTGGCCAGGTGTCTTGTGTTCTCTCTCTCTGAGTACGCGCCCAAGCTGGCT 1503
Qy 2260 TTTCTTGTGGCCAGAGTATTACAGAGAGCCAAATCTGTCTCACTGTCAAAACCGCTTTAC 2319
Db 1504 TTTCTTGTGGCCAGAGTATTACAGAGAGCCAAATCTGTCTCACTGTCAAAACCGCTTTAC 1563
Qy 2320 TACCTCTAA 2328
Db 1564 TACCTCTAA 1572

RESULT 3

ABZ19511
ID ABZ19511 standard; cDNA; 367 BP.
XX
AC ABZ19511;
XX
DT 23-JAN-2003 (first entry)
XX
DE Group III cDNA cancer related clone SEQ ID NO:1937.
KW Human; cancer; tumour; therapy; diagnosis; CT antigen; CP antigen;
KW immune response; virology; immunology; microbiology; molecular biology;
KW recombinant DNA technology; gene; ss.
XX.

OS Homo sapiens.
XX WO200278516-A2.
XX
XX PD 10-OCT-2002.
XX
XX PF 28-MAR-2002; 2002WO-US10421.
XX
XX PR 30-MAR-2001; 2001US-280255P.
XX
XX PR 28-AUG-2001; 2001US-315563P.
XX
XX PR 09-JAN-2002; 2002US-347313P.
XX
XX (CORI-) CORIXA CORP.
XX
XX Wang T, Wang S, Bangur CS, Gaiger A;
XX WPI; 2003-058387/05.
XX
XX PT New immunogenic polynucleotides or polypeptides useful for diagnosing,
XX preventing and treating cancer expressing CT or CP mRNA antigens, and
XX in virology, immunology, microbiology, molecular biology and
XX recombinant DNA techniques
XX
XX PS Claim 1; SEQ ID 1937; 207pp; English.
XX
XX CC ABQ17575 to ABQ20506 represent isolated polynucleotide (I) sequences, and
XX ABP5446 to ABP54472 represent protein (II) sequences, from the present
XX invention. (I) and (II) have cytostatic activity and can be used in gene
XX therapy and vaccines. (I), (II), antibodies and compositions from the
XX present invention are useful for diagnosing, preventing and treating
XX cancer, which expresses CT or CP mRNA antigens. They are useful for
XX stimulating immune response. They can also be useful in virology,
XX immunology, microbiology, molecular biology and recombinant DNA
XX techniques.
XX
XX CC N.B. The sequence data for this patent did not form part of the printed
XX specification, but was obtained in electronic format directly from WIPO
XX at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ Sequence 367 BP; 118 A; 77 C; 91 G; 81 T; 0 other;
Query Match 13.6%; Score 316; DB 25; Length 367;
Best Local Similarity 99.7%; Pred. No. 1.9e-146;
Matches 366; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
Qy 1291 AAGTTACACACCCATGGGCATGCAATATGAGAAAAGCAATATGATTGAAGTGGATGAC 1350
Db 1 AAGTTACACACCCATGGGCATGCAATATGAGAAAAGCAATATGATTGAAGTGGATGAC 60
Qy 1351 AGAAGTGAAGCTTAAAGTCTTACAGCAAAAGTCTACAGACACCCAGATAGTT 1410
Db 61 AGAAGTGAAGCTTAAAGTCTTACAGCAAAAGTCTACAGACACCCAGATAGTT 120
Qy 1411 GTCTGTCTGTGTCAAGTAAATCGAAGGACAAATACGATCTATTAAAAATACCTGTGT 1470
Db 121 GTCTGTCTGTGTCAAGTAAATCGAAGGACAAATACGATCTATTAAAAATACCTGTGT 180
Qy 1471 ACAGATTGCCCTTACCCCAAGTCTAGTGTGTGGCCGCAACCTTAGCACAACAGCAACT 1530
Db 181 ACAGATTGCCCTTACCCCAAGTCTAGTGTGTGGCCGCAACCTTAGCACAACAGCAACT 240
Qy 1531 GTCATGGCCATTGTCTACAAAGATTGCCCTACAGATGAACCTCAAGATGGGAGAGAGCTC 1590
Db 241 GTCATGGCCATTGTCTACAAAGATTGCCCTACAGATGAACCTCAAGATGGGAGAGAGCTC 300
Qy 1591 TGGAGGGTGGACATCCCTTGAAGCTCGTGTATGATCGTTGGCATTCGATTGTACCATGAC 1650
Db 301 TGGAGGGTGGACATCCCTTGAAGCTCGTGTATGATCGTTGGCATTCGATTGTACCATGAC 360
Qy 1651 ATGACAG 1657
Db 361 ATGACAG 367

```
RESULT 4
AAZ13404
ID AAZ13404 standard; cDNA; 300 BP.
XX
AC AAZ13404;
XX
DT 12-OCT-1999 (first entry)
DE Human gene expression product cDNA sequence SEQ ID NO:873.
XX
XX Human; gene; gene expression product; diagnosis; therapy; probe;
KW detection; mapping; tissue typing; profiling; forensic; cancer;
KW genetic analysis; colorectal cancer; breast cancer; lung cancer; ss.
XX
OS Homo sapiens.
XX
XX WO9938972-A2.
XX
XX 05-AUG-1999.
XX
XX 28-JAN-1999; 99WO-US01619.
XX
XX 03-APR-1998; 98US-0080666.
XX
XX 28-JAN-1998; 98US-0072910.
XX
XX 24-FEB-1998; 98US-0075954.
XX
XX 31-MAR-1998; 98US-0080114.
XX
XX 03-APR-1998; 98US-0080515.
XX
XX (CHIR -) CHIRON CORP.
XX
XX (HYSE -) HYSEQ INC.
XX
XX Crkvenjakov R, Dickson M, Drmanac R, Drmanac S;
XX Escobedo J, Garcia PD, Garcia V, Giese K, Innis MA;
XX James WL, Kassam A, Kennedy GC, Kita D, Labat I;
XX Lamson G, Leshkowitz D, Pot D, Randazzo F, Reinhard C;
XX Stache-Crain B, Sudduth-Klinger J, Williams LT;
XX WPI; 1999-494092/41.
XX
XX Novel human genes and their expression products which are
XX differentially expressed in different cell types
XX
XX Claim 1; Page 862-863; 2479pp; English.
XX
XX The present invention describes a library of human polynucleotides
XX comprising the sequences given in AAZ12532 to AAZ17779. Also described is
XX a method of detecting differentially expressed genes correlated with the
XX cancerous state of a mammalian cell, comprising detecting at least one
XX differentially expressed gene product in a test sample from a cell
XX suspected of being cancerous, where the gene product is encoded by one
XX of the 548 polynucleotide sequences given in AAZ12532 to AAZ1779. The
XX polynucleotides can be used as a source of primers and probes, which can
XX be used for a variety of purpose, e.g. detection of expression levels,
XX mapping, tissue typing or profiling, forensics, genetic analysis and
XX detection of polymorphisms. Polypeptides encoded by the polynucleotides
XX can be used for raising antibodies for experimental, diagnostic and
XX therapeutic purposes. The polynucleotides may also be used to construct
XX arrays for diagnostics (which may be used to determine function of an
XX encoded protein); and to detect differences in expression levels between
XX two cells (e.g. to identify abnormal or diseased tissue in a human, to
XX identify a genetic predisposition or susceptibility to a disease such as
XX cancer). The polynucleotides of the invention are especially used in the
XX diagnosis, prognosis and management of colorectal cancer, breast cancer,
XX and lung cancer. The polynucleotides can also be used to screen for
XX peptide analogues and antagonists.
XX
XX Sequence 300 BP; 75 A; 70 C; 91 G; 64 T; 0 other;
XX
XX Query Match 12.2%; Score 284; DB 20; Length 300;
XX Best Local Similarity 100.0%; Pred. No. 1.5e-130;
XX Matches 284; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
XX
XX 1485 CCCAAGTCAGTGTGTGGTGGCCGACCTTAGGCAACAGCAACTGTGTCATGGCCATTGC 1544
XX
```

```
Db 1 CCCAAGTCAGTGTGTGTGGTGGCCGACCTTAGGCAACAGCAACTGTGTCATGGCCATTGC 60
Qy 1545 TACAAAGATTCCCTACAGATGAAGTCAAGATGGAGAGAGCTCTCTGAGGGTGAACAT 1604
Db 61 TACAAAGATTCCCTACAGATGAAGTCAAGATGGAGAGAGCTCTCTGAGGGTGAACAT 120
Qy 1605 CCCCTGAAGCTCGTGATGATCGTTGGCATCGATTGTTTACCATGACATGACAGCTGGGCG 1664
Db 121 CCCCTGAAGCTCGTGATGATCGTTGGCATCGATTGTTTACCATGACATGACAGCTGGGCG 180
Qy 1665 GAGTCAATCGCAGGATTGTTGCCAGCATCAATCAAGGATGACCCCTGTTCTCAG 1724
Db 181 GAGTCAATCGCAGGATTGTTGCCAGCATCAATCAAGGATGACCCCTGTTCTCAG 240
Qy 1725 CTGCATATTTTCCAGGATAGAGGACAGGAGCTGTGTAGATGGGTCA 1768
Db 241 CTGCATATTTTCCAGGATAGAGGACAGGAGCTGTGTAGATGGGTCA 284

RESULT 5
AAF21880
ID AAF21880 standard; DNA; 423 BP.
XX
AC AAF21880;
XX
XX 27-MAR-2001 (first entry)
XX
XX Human breast and ovarian cancer associated antigen gene SEQ ID 267.
XX
XX Human; breast cancer; ovarian cancer; cytostatic; immunosuppressive;
XX neurotropic; neuroprotective; antiviral; antiallergic; hepatotropic;
XX antidiabetic; antiinflammatory; antitumor; antiparasitic; anticonvulsant;
XX antibacterial; antifungal; antiparasitic; cardiac; immune disorder;
XX Addison's disease; allergy; autoimmune haemolytic anaemia;
XX autoimmune thyroiditis; diabetes mellitus; Crohn's disease;
XX multiple sclerosis; rheumatoid arthritis; ulcerative colitis;
XX cardiovascular disorder; wound healing; neurological disease; ds.
XX
XX Homo sapiens.
XX
XX WO200055173-A1.
XX
XX 21-SEP-2000.
XX
XX 08-MAR-2000; 2000WO-US05881.
XX
XX 12-MAR-1999; 99US-0124270.
XX
XX (HUMA-) HUMAN GENOME SCI INC.
XX
XX Rosen CA, Ruben SM;
XX WPI; 2000-611515/S9.
XX
XX P-PSDB; AAB58977.
XX
XX New human breast and ovarian cancer associated gene sequences and the
XX polypeptides encoded by these genes, useful in the prevention,
XX treatment and diagnosis of cancer, immune disorders, cardiovascular
XX disorders and neurological diseases -
XX
XX Claim 1; Page 686; 1299pp; English.
XX
XX Sequences AAF21614 - AAF22031 represent DNA sequences encoding human
XX proteins AAB58711 - AAB59128. The DNA and protein sequences are
XX associated with breast and ovarian cancer. Included in the invention are
XX sequences AAF22032 - AAF22040 and AAB59129 which are used in the
XX isolation and characterisation of the DNA and protein sequences of the
XX invention. The breast and ovarian cancer associated DNA, protein, agonist
XX or antagonist sequences exhibit cytostatic; immunosuppressive;
XX neurotropic; neuroprotective; antiviral; antiallergic; hepatotropic;
XX antidiabetic; antiinflammatory; antitumor; antiparasitic; anticonvulsant;
XX antibacterial; antifungal; antiparasitic and cardiac activity. The
```



```
OS Homo sapiens.
XX WO200278516-A2.
XX 10-OCT-2002.
XX
XX 28-MAR-2002; 2002WO-US10421.
XX
XX 30-MAR-2001; 2001US-280255P.
XX
XX 28-AUG-2001; 2001US-315563P.
XX
XX 09-JAN-2002; 2002US-347313P.
XX
XX (CORI-) CORIXA CORP.
XX
XX Wang T, Wang S, Bangur CS, Gaiger A;
XX WPI; 2003-058387/05.
XX
XX New immunogenic polynucleotides or polypeptides useful for diagnosing,
XX preventing and treating cancer expressing CT or CP mRNA antigens, and
XX PT in virology, immunology, microbiology, molecular biology and
XX PT recombinant DNA techniques -
XX
XX Claim 1; SEQ ID 1528; 207pp; English.
XX
XX ABQ17575 to ABQ20506 represent isolated polynucleotide (I) sequences, and
XX CC ABP54446 to ABP54472 represent protein (II) sequences, from the present
XX CC invention. (I) and (II) have cytostatic activity and can be used in gene
XX CC therapy and vaccines. (I), (II), antibodies and compositions from the
XX CC present invention are useful for diagnosing, preventing and treating
XX CC cancer, which expresses CT or CP mRNA antigens. They are useful for
XX CC stimulating immune response. They can also be useful in virology,
XX CC immunology, microbiology, molecular biology and recombinant DNA
XX CC techniques.
XX CC N.B. The sequence data for this patent did not form part of the printed
XX CC specification, but was obtained in electronic format directly from WIPO
XX CC at ftp.wipo.int/pub/published_pct_sequences.
XX
XX Sequence 209 BP; 48 A; 47 C; 70 G; 44 T; 0 other;
XX
XX Query Match 6.8%; Score 158; DB 25; Length 209;
XX Best Local Similarity 99.5%; Pred. No. 7e-68;
XX Matches 208; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
XX
XX QY 1658 CTGGCGGAGTCAATCGCAGGATTGTTCAGCATCAATGAGGAGTACCGCGTGGT 1717
XX DB 1 CTGGCGGAGTCAATCGCAGGATTGTTCAGCATCAATGAGGAGTACCGCGTGGT 60
XX
XX QY 1718 TCTCACGCTGCATATTTTCAGGATAGAGGACAGGAGCTGTAGATGGGCTCAAAGTCTGCC 1777
XX DB 61 TCTCACGCTGCATATTTTCAGGATAGAGGACAGGAGCTGTAGATGGGCTCAAAGTCTGCC 120
XX
XX QY 1778 TGCAAGCGGCTCTGAGGCTTGGAAATAGCTCAATAGTACATGCCAGCCGATCATCG 1837
XX DB 121 TGCAAGCGGCTCTGAGGCTTGGAAATAGCTCAATAGTACATGCCAGCCGATCATCG 180
XX
XX QY 1838 TGTACCGGATGGCGTAGGACAGCCGAG 1866
XX DB 181 TGTACCGGATGGCGTAGGACAGCCGAG 209
XX
XX RESULT 8
XX ABN42800
XX ID ABN42800 standard; DNA; 60 BP.
XX
XX AC ABN42800;
XX
XX 15-JUL-2002 (first entry)
XX
XX Human spliced transcript detection oligonucleotide SEQ ID NO:1548.
XX
XX Human; mouse; rat; splice transcript; detection; RNA transcript;
XX KW splice variant; transcriptome; oligonucleotide library; ss.
```

```
XX Homo sapiens.
XX OS WO200210449-A2.
XX
XX 07-FEB-2002.
XX
XX 20-JUL-2001; 2001WO-IB01903.
XX
XX 28-JUL-2000; 2000US-221607P.
XX
XX 02-MAY-2001; 2001US-287724P.
XX
XX (COMP-) COMPUGEN INC.
XX
XX Shoshan A, Wasserman A, Mintz E, Mintz L, Paigler S;
XX WPI; 2002-257383/30.
XX
XX New oligonucleotide libraries comprising oligonucleotides which
XX PT selectively hybridize to mRNAs transcribed from a transcription unit of
XX PT a genome, useful for detecting tissue-, pathology-, and
XX PT developmental-specific genes -
XX
XX Example 1; SEQ ID 15548; 47pp; English.
XX
XX The present invention describes oligonucleotide libraries for detecting
XX CC messenger RNAs that populate a (sub-)transcriptome, where the
XX CC (sub-)transcriptome comprises messenger RNAs transcribed from multiple
XX CC transcription units that populate a genome. The library comprises
XX CC several oligonucleotides, each capable of hybridising selectively to a
XX CC set of messenger RNAs transcribed from a given transcription unit of
XX CC the genome, which encodes one or more messenger RNA splice variants.
XX CC The oligonucleotide libraries are useful for detecting mRNAs from a
XX CC biological sample, in expression profiling studies, in qualitatively or
XX CC quantitatively characterising the corresponding transcriptome, and in
XX CC detecting RNA transcripts and splice variants of human or animal
XX CC transcriptomes. The libraries may also be used as specialised mini-
XX CC libraries to detect transcripts of a sub-transcriptome under a
XX CC particular biological or pathological state, and so allowing the
XX CC detection of tissue- and pathology-specific genes such as those genes
XX CC only expressed in specific tissue under a specific pathological
XX CC condition; to detect developmental specific genes; and to detect RNA
XX CC transcripts and splice variants of a transcriptome of a patient suffering
XX CC from a particular disorder. ABN27253 to ABN59589 represent
XX CC oligonucleotide sequences from rats, humans and mice, which are used in
XX CC the exemplification of the present invention.
XX CC N.B. The sequence data for this patent did not form part of the printed
XX CC specification, but was obtained in electronic format directly from WIPO
XX CC at ftp.wipo.int/pub/published_pct_sequences.
XX
XX Sequence 60 BP; 23 A; 9 C; 17 G; 11 T; 0 other;
XX
XX Query Match 2.6%; Score 60; DB 24; Length 60;
XX Best Local Similarity 100.0%; Pred. No. 3.8e-19;
XX Matches 60; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
XX
XX QY 1924 GGTAGAGTTACACCTAGACTAACGGTAATGTGTGTAAGAAAGAGTGAACACGAGA 1983
XX DB 1 GGTAGAGTTACACCTAGACTAACGGTAATGTGTGTAAGAAAGAGTGAACACGAGA 60
XX
XX RESULT 9
XX AAS25525/C
XX ID AAS25525 standard; cDNA; 460 BP.
XX
XX AC AAS25525;
XX
XX 07-NOV-2001 (first entry)
XX
XX Human ovarian PCR-subtracted cDNA library clone #1610.
XX
XX Immunogenic protein; cancer; ovarian tumour; T-cell stimulation; ss;
XX KW gene therapy; cytostatic; T-cell expansion; nucleic acid hybridisation;
```

```
KW primer; probe.
XX OS
XX Homo sapiens.
XX PN WO200157207-A2.
XX PD 09-AUG-2001.
XX PF 05-FEB-2001; 2001WO-US03733.
XX PR 04-FEB-2000; 2000US-0180403.
XX PR 28-MAR-2000; 2000US-0192745.
XX PA (CORI-) CORIXA CORP.
XX PI
XX PI Algate PA, Mannion J;
XX DR WPI; 2001-488879/53.
XX
XX New polynucleotides encoding ovarian tumour proteins, useful for
PT treating ovarian cancer, and as probes, primers, and markers of cancer
PT progression
XX
XX Example 1; page 374; 378pp; English.
XX
XX The invention comprises compositions used for the therapy and diagnosis
CC of ovarian cancer. The compositions comprise one or more ovarian tumour
CC proteins, their associated polynucleotides, or immunogenic portions of
CC the proteins. The ovarian tumour polynucleotides and polypeptides are
CC useful for stimulating and/or expanding T cells specific for a tumour
CC protein. They are also useful for inhibiting the development of cancer in
CC a patient with an ovarian tumour DNA or protein by incubating isolated
CC T-cells allowing them to proliferate, and administering to the patient.
CC The sequences can be used as markers for cancer, for example, to monitor
CC ovarian cancer progression. Probes and primers are useful in nucleic acid
CC hybridisation, in detecting the presence of complementary sequences in a
CC given sample, for preparing mutant species and for preparing other
CC genetic constructions. Sequences AAS23820-AAS25231 and AAS25328-AAS25549
CC represent human ovarian tumour protein cDNA clones.
XX
XX Sequence 460 BP; 188 A; 78 C; 69 G; 125 T; 0 other;
SQ
```

```
Query Match 2.5%; Score 58; DB 22; Length 460;
Best Local Similarity 100.0%; Pred. No. 3.7e-18;
Matches 58; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2271 CCAGAGTATTCACAGAGAGCCAAATCTGTCACTGTCAAAACCGCTTTACTACTCTAA 2328
Db 460 CCAGAGTATTCACAGAGAGCCAAATCTGTCACTGTCAAAACCGCTTTACTACTCTAA 403

RESULT 10
AAA07587
ID AAA07587 standard; DNA; 4064 BP.
XX
XX AAA07587;
AC
XX
XX 29-AUG-2000 (first entry)
XX
XX Mouse piwi gene, designated miwi.
XX
XX Piwi family protein; piwi; miwi; hiwi; gene therapy; tissue dystrophy;
KW anaemia; immunodeficiency; male infertility; mouse; ds.
XX
XX Mus sp.
OS
XX
XX Key Location/Qualifiers
XX CDS 191..2779
XX /*tag= a
XX /product= miwi
XX /transl_except= (pos:450..452; aa:Xaa)
XX /transl_except= (pos:1337..1339; aa:Xaa)
XX /transl_except= (pos:2636..2638; aa:Xaa)
```

```
FT
FT
XX
XX
XX PN WO200032039-A1.
XX PD 08-JUN-2000.
XX PF 03-DEC-1999; 99WO-US28764.
XX PR 04-DEC-1998; 98US-0110901.
XX PA (UYDU-) UNIV DUKE.
XX PI
XX PI Lin H;
XX DR WPI; 2000-412085/35.
XX DR P-PSDB; AAY90234.
XX
XX Piwi family nucleic acids, polypeptides, and antibodies, useful in gene
PT therapy of diseases such as cancer and in various research and
PT diagnostic applications
XX
XX Claim 19; Page 180-185; 201pp; English.
XX
XX This sequence encodes the mouse piwi family protein, designated
CC miwi. The piwi family nucleic acids and polypeptides are used in gene
CC therapy of diseases such as cancer and also in various research and
CC diagnostic applications. The sequences can also be used to treat
CC tissue dystrophy, anaemia, immunodeficiency, and male infertility.
XX
XX Sequence 4064 BP; 1114 A; 978 C; 1077 G; 890 T; 5 other;
SQ
```

```
Query Match 1.5%; Score 35; DB 21; Length 4064;
Best Local Similarity 100.0%; Pred. No. 9.9e-07;
Matches 35; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1687 GCCAGCATCAATGAAGGGATGACCCGCTGGTCTC 1721
Db 2138 GCCAGCATCAATGAAGGGATGACCCGCTGGTCTC 2172

RESULT 11
ABK44483
ID ABK44483 standard; cDNA; 501 BP.
XX
XX ABK44483;
AC
XX
XX 05-JUN-2002 (first entry)
XX
XX cDNA encoding colon tumour protein, SEQ ID No 34.
XX
XX Human; colon tumour; vaccine; colon cancer; immunogenic;
KW immunotherapy; gene; ss.
XX
XX Homo sapiens.
OS
XX
XX WO200212328-A2.
XX
XX 14-FEB-2002.
XX
XX 31-JUL-2001; 2001WO-US24218.
XX
XX 03-AUG-2000; 2000US-223283P.
XX
XX 28-MAR-2001; 2001US-279763P.
XX
XX 29-JUN-2001; 2001US-302051P.
XX
XX (CORI-) CORIXA CORP.
XX
XX King GE, Meagher MJ, Xu J, Secrist H;
XX
XX WPI; 2002-241739/29.
XX
```

PT New colon cancer polypeptides and polynucleotides, useful as vaccines,
PT for diagnosing, preventing, and treating colon cancer, and as markers
PT for the progression of cancer -
XX
PS Claim 1; SEQ ID No 34; 147pp; English.
XX
CC The invention relates to polynucleotides encoding colon tumour proteins.
CC The polynucleotides and encoded polypeptides are useful in pharmaceutical
CC compositions, such as vaccines, for the diagnosis, prevention, and
CC treatment of colon cancer. Polynucleotide sequences may be used as
CC hybridisation probes or primers, and in the design and preparation of
CC ribozyme molecules for inhibiting expression of tumour polypeptides and
CC proteins in tumour cells. The compositions are useful for stimulating an
CC immune response against cancer, particularly for the immunotherapy of
CC colon cancer, and as markers for the progression of cancer.
CC ABK4450-ABK46237 represent coding sequences of human colon tumour
CC proteins of the invention.
CC Note: With the exception of SEQ ID No 1 and 2, the sequence data
CC for this patent did not form part of the printed specification but was
CC supplied by the European Patent Office.
XX
SQ Sequence 501 BP; 134 A; 81 C; 93 G; 193 T; 0 other;
Query Match 1.5%; Score 34; DB 24; Length 501;
Best Local Similarity 100.0%; Pred. No. 3.2e-06;
Matches 34; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 2271 CCAGATATTTCACAGAGAGCCAAATCTGTCACGTG 2304
Db 1 CCAGATATTTCACAGAGAGCCAAATCTGTCACGTG 34
RESULT 12
ABX15293/C
ID ABX15293 standard; DNA; 22 BP.
XX
AC ABX15293;
XX
DT 27-MAR-2003 (first entry)
XX
DE ASM 698-13 B_1 PCR primer ASM75-3' T.
XX
KW ss: allele frequency; proofreading polymerase; gene typing; PCR;
KW karyotyping; genotyping; DNA family planning; diagnostics; primer;
KW prenatal testing; paternal determination; pharmacogenetics;
KW forensic analysis; ASM 698-13 B_1.
XX
OS Unidentified.
XX
PH Key Location/Qualifiers
FT modified_base 22 /*tag= a
FT FT /mod base= OTHER
FT FT /note= "T is labeled with a ROX (not defined) moiety"
XX
PN US2002142336-A1.
XX
XX 03-OCT-2002.
XX
XX 01-FEB-2002; 2002US-0061961.
XX
XX 02-FEB-2001; 2001US-266035P.
XX
XX (GENO-) GENOME THERAPEUTICS CORP.
XX
XX Smith DR, Thomann H, Cahill P;
XX
XX WPI; 2003-182355/18.
XX
XX Detecting the presence or absence of a first nucleotide at position
PT within a strand of DNA, useful in gene typing, genotyping, disease
PT diagnostics, prenatal testing, paternal determination, pharmacogenetics
PT and forensic analysis -

XX
PS Example 2; Page 13; 40pp; English.
XX
CC The invention relates to detecting the presence or absence of a first
CC nucleotide, at a position within a strand of DNA, comprising forming an
CC admixture of a primer and the strand of DNA, with or without a mixture of
CC labeled dideoxynucleotides and forming a hybridisation product, where the
CC primer comprises an electrophoretic tag (e-tag) group or fluorescent
CC label opposite the first nucleotide. Also included is a method for
CC determining allele frequency at a first nucleotide position within a
CC strand of DNA in a sample comprising: (a) providing a first primer
CC comprising a sequence of DNA which hybridises with the strand of DNA
CC adjacent to the first nucleotide position, and having a second nucleotide
CC opposite the first nucleotide position, the second nucleotide hybridising to the
CC with a first detectable label, the second nucleotide hybridising to the
CC first nucleotide in the event the second nucleotide is complementary to the
CC first nucleotide and the second nucleotide not hybridising to the
CC first nucleotide in the event the second nucleotide is not complementary;
CC (b) providing a second primer, the second primer comprising a sequence
CC of DNA which hybridises with the strand of DNA adjacent to the first
CC nucleotide position, and having a third nucleotide opposite the first
CC nucleotide position, the third nucleotide associated with a second
CC detectable label, and hybridising or not hybridising to the first
CC nucleotide in the event the third nucleotide is complementary or not
CC complementary to the first nucleotide; (c) forming an admixture of the
CC first and second primers and the strand of the DNA strand to form a
CC hybridisation product; (d) applying a proofreading polymerase to the
CC hybridisation product under conditions in which the second and third
CC nucleotide is preferentially excised in the event the second and third
CC nucleotide is not hybridised to the first nucleotide, and in which the
CC second and third nucleotide is preferentially incorporated into an
CC extension product in the event the second and third nucleotide is
CC hybridised to the first nucleotide; and (e) monitoring the sample for the
CC presence of a first or a second label in association with the extension
CC product, where the ratio of the first and second label is indicative of
CC allele frequency at the first nucleotide position within a strand of DNA
CC in a sample. The methods of the present invention are useful in gene
CC typing, karyotyping, genotyping, DNA family planning, diagnostics,
CC prenatal testing, paternal determination, pharmacogenetics and forensic
CC analysis. the present sequence is a labeled PCR primer used to
CC investigate the ASM 698-13 B_1 (not defined) gene region using the method
CC of the invention.
XX
SQ Sequence 22 BP; 5 A; 5 C; 8 G; 4 T; 0 other;
Query Match 0.9%; Score 21; DB 25; Length 22;
Best Local Similarity 100.0%; Pred. No. 9.6;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 871 CCAGGGCGCTGCCATGCTCATT 891
Db 21 CCAGGGCGCTGCCATGCTCATT 1
RESULT 13
AAV21209
ID AAV21209 standard; DNA; 1664976 BP.
XX
AC AAV21209;
XX
XX 10-NOV-1998 (first entry)
XX
XX Methanococcus jannaschii circular chromosome.
XX
XX Methanococcus jannaschii; methanogenic archaeon; circular chromosome;
KW genome; autotrophic; extrachromosomal element; identification; ds.
XX
XX Methanococcus jannaschii.
OS
XX
XX WO9807830-A2.
XX
PD 26-FEB-1998.
XX

ABN23657/c
ID ABN23657 standard; cDNA; 302 BP.
XX AC ABN23657;
XX DT 24-JUN-2002 (first entry)
XX DE Human ORFX polynucleotide sequence SEQ ID NO:15791.
XX KW Human; open reading frame; ORFX; gene therapy; cancer; cirrhosis;
KW hyperproliferative disorder; psoriasis; benign tumour; haemorrhage;
KW degenerative disorder; osteoarthritis; neurodegenerative disorder;
KW cardiovascular disease; diabetes mellitus; systemic lupus erythematosus;
KW hypertension; hypothyroidism; cholesterol ester storage disease;
KW immune deficiency; immune disorder; infectious disease;
KW autoimmune disorder; rheumatoid arthritis; autoimmune thyroiditis;
KW myasthenia gravis; Gene; ss.
XX OS Homo sapiens.
XX PN WO200192523-A2.
XX PD 06-DEC-2001.
XX PF 29-MAY-2001; 2001WO-US10836.
XX PR 30-MAY-2000; 2000US-206132P.
XX PR 29-AUG-2000; 2000US-228716P.
XX PA (CURA-) CURAGEN CORP.
XX PI Shimkets RA, Leach MD;
XX DR WPI; 2002-106308/14.
XX DR P-PSDB; ABP07905.

Novel human polypeptides and polynucleotides useful for diagnosing,
preventing and treating cardiovascular disease, neurodegenerative,
hyperproliferative disorders and autoimmune disorders -
Disclosure; SEQ ID 15791; 1037pp; English.

The present invention describes substantially purified human proteins
(referred to as open reading frame, ORFX, where X is 1-11491 (see Table 1
in the specification). ABN15762 to ABN27252 encode the human ORFX
proteins given in ABP00010 to ABP11500. ORFX proteins are useful for
treating or preventing a pathology associated with an ORFX-associated
disorder in humans, and in the manufacture of a medicament for treating a
syndrome associated with ORFX-associated disorder. ORFX polynucleotide
sequences can be used in gene therapy. ORFX sequences can be used in the
treatment of cancer, hyperproliferative disorders, cirrhosis of liver,
psoriasis, benign tumours, keloid, degenerative disorders, haemorrhage,
osteoarthritis, neurodegenerative disorders, disorders related to organ
transplantation, cardiovascular diseases, diabetes mellitus, systemic
lupus erythematosus, hypertension, hypothyroidism, cholesterol ester
storage disease, various immune deficiencies and disorders, infectious
diseases, autoimmune disorders such as multiple sclerosis, rheumatoid
arthritis, autoimmune thyroiditis, myasthenia gravis, graft-versus-host
disease and autoimmune inflammatory eye disease. ORFX proteins are also
useful for treating burns, incisions, ulcers, for treating osteoporosis,
bone degenerative disorders, or periodontal disease, and for gut
protection or regeneration and treatment of lung or liver fibrosis,
reperfusion injury in various tissues and conditions resulting from
systemic cytokine damage.
N.B. The sequence data for this patent did not form part of the printed
specification, but was obtained in electronic format directly from WIPO
at ftp.wipo.int/pub/published_pct_sequences.

Sequence 302 BP; 90 A; 54 C; 58 G; 96 T; 4 other;

Query Match 0.8%; Score 19; DB 24; Length 302;
Best Local Similarity 100.0%; Pred.No. 92;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 635 AAGTTTCCAAAGCACTAAT 653
Db 165 AAGTTTCCAAAGCACTAAT 147

Search completed: December 6, 2003, 18:22:26
Job time : 622 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: December 10, 2003, 17:32:26 ; Search time 42 Seconds

(without alignments)
4761.685 Million cell updates/sec

Title: US-10-043-774B-2

Perfect score: 775

Sequence: 1 MIFGVTQNLQDHVKESKTG.....VQSTHREPNLNRLYYL 775

Scoring table:

Gapop 60.0 , Gapext 60.0

Searched: 830525 seqs, 258052604 residues

Word size : 0

Total number of hits satisfying chosen parameters: 830525

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database :

SPTREMBL.23.*
1: sp_archaea.*
2: sp_bacteria.*
3: sp_fungi.*
4: sp_human.*
5: sp_invertebrate.*
6: sp_mammal.*
7: sp_mhc.*
8: sp_organelle.*
9: sp_phage.*
10: sp_plant.*
11: sp_rodent.*
12: sp_virus.*
13: sp_vertebrate.*
14: sp_unclassified.*
15: sp_rvirus.*
16: sp_bacteriap.*
17: sp_archaeap.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Match	Query	ID	Description
1	775	100.0	775	4 Q96JD5	Q96jd5 homo sapien
2	772	99.6	861	4 Q96J94	Q96j94 homo sapien
3	645	83.2	829	4 Q8NA60	Q8na60 homo sapien
4	570	73.5	861	4 Q95404	Q95404 homo sapien
5	469	60.5	861	4 Q8TBY5	Q8tby5 homo sapien
6	123	15.9	862	11 Q9JMB7	Q9jmb7 mus musculu
7	21	2.7	858	13 Q8UVX0	Q8uvx0 brachydanio
8	16	2.1	808	5 Q9GPA7	Q9gpa7 strongyloce
9	16	2.1	854	5 Q9GPA8	Q9gpa8 strongyloce
10	15	1.9	722	5 Q17567	Q17567 caenorhabdi
11	15	1.9	824	5 P90786	P90786 caenorhabdi
12	11	1.4	121	11 Q8CC75	Q8cc75 mus musculu
13	11	1.4	666	4 Q8N9G9	Q8n9g9 homo sapien
14	11	1.4	852	4 Q8N9V8	Q8n9v8 homo sapien
15	11	1.4	852	4 Q8NEH2	Q8neh2 homo sapien
16	11	1.4	878	11 Q8CGT6	Q8cgt6 mus musculu

17	9	1.2	189	5 Q95P85	Q95p85 drosophila
18	9	1.2	866	5 Q76922	Q76922 drosophila
19	9	1.2	1035	5 Q21079	Q21079 caenorhabdi
20	9	1.2	1037	5 Q9TW94	Q9tw94 caenorhabdi
21	8	1.0	371	4 Q9NW28	Q9nw28 homo sapien
22	8	1.0	398	10 Q80872	Q80872 arabidopsis
23	8	1.0	421	5 Q9GPA6	Q9gpa6 strongyloce
24	8	1.0	449	10 Q8S9X7	Q8s9x7 oryza sativ
25	8	1.0	530	4 Q96SW6	Q96sw6 homo sapien
26	8	1.0	580	11 Q99WV6	Q99mw6 mus musculu
27	8	1.0	704	5 Q95XQ7	Q95xq7 caenorhabdi
28	8	1.0	730	2 Q07667	Q07667 enterococu
29	8	1.0	971	11 Q9JMB6	Q9jmb6 mus musculu
30	8	1.0	971	11 Q8CDG1	Q8cdg1 mus musculu
31	8	1.0	973	4 Q8TCS9	Q8tcs9 homo sapien
32	8	1.0	1137	5 Q93250	Q93250 caenorhabdi
33	8	1.0	1498	13 Q8UUM8	Q8uum8 oryzias lat
34	7	0.9	41	2 Q8G96	Q8g96 neisseria p
35	7	0.9	61	16 Q8YZK9	Q8yzk9 anabaena sp
36	7	0.9	68	16 Q8FKT9	Q8fkt9 escherichia
37	7	0.9	77	16 Q9ZJ23	Q9zj23 helicobacte
38	7	0.9	83	17 Q8TSK7	Q8tsk7 methanosarc
39	7	0.9	85	2 Q8S556	Q8s56 pseudomonas
40	7	0.9	85	16 Q9S564	Q9s564 pseudomonas
41	7	0.9	86	8 Q95817	Q95817 hymenolepis
42	7	0.9	86	16 Q8PDT6	Q8pdt6 escherichia
43	7	0.9	90	2 Q8GC58	Q8gc58 escherichia
44	7	0.9	92	2 Q93F09	Q93f09 shigella fl
45	7	0.9	92	16 Q8FG80	Q8fg80 escherichia

ALIGNMENTS

RESULT 1
Q96JD5 PRELIMINARY; PRT; 775 AA.
ID Q96JD5
AC Q96JD5; DT 01-DEC-2001 (TRENBLrel. 19, Created)
DT 01-DEC-2001 (TRENBLrel. 19, Last sequence update)
DT 01-OCT-2002 (TRENBLrel. 22, Last annotation update)
DE HIWI.
OS Homo sapiens (Human)
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Testis;
RA Sharma A.K., Nelson M.C., Brandt J.E., Wessman M., Muhud N.,
RA Weller K.P., Hoffman R.;
RT "Human CD34+ Stem Cells Express the hiwi Gene, a Human Homolog of the
RT Drosophila Gene piwi.";
RL Submitted (MAY-2000) to the EMBL/GenBank/DBSJ databases.
DR EMBL; AF264004; AAK92281.1; -
DR InterPro; IPR003100; PAZ.
DR InterPro; IPR003165; Piwi.
DR Pfam; PF02170; PAZ; 1.
DR Pfam; PF02171; Piwi; 1.
DR PROSITE; PS0821; PAZ; 1.
DR PROSITE; PS0822; PIWI; 1.
SQ SEQUENCE 775 AA; 89484 MW; DF169A2E9EAFD916 CRC64;

Query Match 100.0%; Score 775; DB 4; Length 775;
Best Local Similarity 100.0%; Pred.No. 0;
Matches 775; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 MIFGVTQNLQDHVKESKTGSGGIIVRLSTHFRITSPQWALYQYHIDYNPLMEARRLR 60
DB 1 MIFGVTQNLQDHVKESKTGSGGIIVRLSTHFRITSPQWALYQYHIDYNPLMEARRLR 60
QY 61 SALLPQHEDLICKCHAFDGTILFLPKRQQKQVTEFSTRNGEDVRITITLNEPPTSP 120

Db 61 SALLFQHEDLKCHAFDGTILFLPKRLQKQVTEVSKTRNGEDVRITITLTNELPPTSP 120
 QY 121 TGLQFYNIIFRLLKIMNLQOIGRNYNPNPDIDIPSHRLVIWPGFTTTSILOYENSIMLC 180
 Db 121 TGLQFYNIIFRLLKIMNLQOIGRNYNPNPDIDIPSHRLVIWPGFTTTSILOYENSIMLC 180
 QY 181 TDVSHKVLRSSETVLDPMFNFYHQTTEHKFQEQVSKELIGLVLTCKYNNKTYVDDIDWDQ 240
 Db 181 TDVSHKVLRSSETVLDPMFNFYHQTTEHKFQEQVSKELIGLVLTCKYNNKTYVDDIDWDQ 240
 QY 241 NPKSTFKKADGSGSVSFLEYRKYNOEITDLKQPVLSQPKRRRGGTLPGPAMLIPEL 300
 Db 241 NPKSTFKKADGSGSVSFLEYRKYNOEITDLKQPVLSQPKRRRGGTLPGPAMLIPEL 300
 QY 301 CYLTGLTDMKRNDFNVMKDLAVHTRLTPEQREVGELIDYIHKNNDVQRELDMGLSPD 360
 Db 301 CYLTGLTDMKRNDFNVMKDLAVHTRLTPEQREVGELIDYIHKNNDVQRELDMGLSPD 360
 QY 361 SNLLSPSGRILQTEKHOGGKTFDYNPOFADWSKETRGAPLISVKPLDNWLLIYTRNYYE 420
 Db 361 SNLLSPSGRILQTEKHOGGKTFDYNPOFADWSKETRGAPLISVKPLDNWLLIYTRNYYE 420
 QY 421 AANSLIQLFKVTPAMGMQMRKAIMIEVDRTTEAYLRVLQKVTADTQIVVCLLSNRKD 480
 Db 421 AANSLIQLFKVTPAMGMQMRKAIMIEVDRTTEAYLRVLQKVTADTQIVVCLLSNRKD 480
 QY 481 KYDAIKKYLCTDPTSPQCVARTLGKQQTWMAIATKIALQMNCKMGELWRVDIPLKLV 540
 Db 481 KYDAIKKYLCTDPTSPQCVARTLGKQQTWMAIATKIALQMNCKMGELWRVDIPLKLV 540
 QY 541 MIVGIDCYHDMTAGRSIAGFVASINEGTRWFSRCIFQDRGOELVDGLKVCLOALRAW 600
 Db 541 MIVGIDCYHDMTAGRSIAGFVASINEGTRWFSRCIFQDRGOELVDGLKVCLOALRAW 600
 QY 601 NSCNEMPSRIIYVRDGVGGQKLTIVNYEVPQFLDCLKSIIGRGNPRLTVIIVKRVNT 660
 Db 601 NSCNEMPSRIIYVRDGVGGQKLTIVNYEVPQFLDCLKSIIGRGNPRLTVIIVKRVNT 660
 QY 661 RPPAQSGRQLNPLPGTVIDVEVTRPEWYDFFIVSQAVRSVSGVSPHYNVIYDNSGLKPD 720
 Db 661 RPPAQSGRQLNPLPGTVIDVEVTRPEWYDFFIVSQAVRSVSGVSPHYNVIYDNSGLKPD 720
 QY 721 HIQRLTYKLCHIIYNNPVGIRVPAPCOYAHKLAFLVGSQSIHREPNSLSNRLYYL 775
 Db 721 HIQRLTYKLCHIIYNNPVGIRVPAPCOYAHKLAFLVGSQSIHREPNSLSNRLYYL 775

RESULT 2

Q96J94 PRELIMINARY; PRT; 861 AA.
 AC Q96J94;
 DT 01-DEC-2001 (TrEMBLrel. 19, Created)
 DT 01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
 DT 01-OCT-2002 (TrEMBLrel. 22, Last annotation update)
 DE PIWI protein.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 OC NCBI_TaxID=9606;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC TISSUE=Testis;
 RA Sha J.H.;
 RT "Cloning and identification of human piwi protein related to testis development."
 RL Submitted (MAY-2001) to the EMBL/GenBank/DBJ databases.
 DR EMBL: AF387507; AAK69348.1; -
 DR InterPro; IPR003100; PAZ.
 DR InterPro; IPR003165; PAZ.
 DR Pfam; PF02170; PAZ; 1.
 DR Pfam; PF02171; Piwi; 1.
 DR PROSITE; PS0821; PAZ; 1.
 DR PROSITE; PS0822; PIWI; 1.

SQ SEQUENCE 861 AA; 98603 MW; 58D7F6C7321DEFA4 CRC64;
 Query Match 99.6%; Score 772; DB 4; Length 861;
 Best Local Similarity 100.0%; Pred.No. 0;
 Matches 772; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 4 GVNTRQNLHDHYKESKSGSGLIIVRLSTNNHPLTSRPOWALYQYHIDYNPLMEARLRSL 63
 Db 90 GVNTRQNLHDHYKESKSGSGLIIVRLSTNNHPLTSRPOWALYQYHIDYNPLMEARLRSL 149
 QY 64 LFQHEDLGKCHAFDGTILFLPKRLQKQVTEVSKTRNGEDVRITITLTNELPPTSPCL 123
 Db 150 LFQHEDLGKCHAFDGTILFLPKRLQKQVTEVSKTRNGEDVRITITLTNELPPTSPCL 209
 QY 124 QFYNIIFRLLKIMNLQOIGRNYNPNPDIDIPSHRLVIWPGFTTTSILOYENSIMLCIDV 183
 Db 210 QFYNIIFRLLKIMNLQOIGRNYNPNPDIDIPSHRLVIWPGFTTTSILOYENSIMLCIDV 269
 QY 184 SHKVLRSSETVLDPMFNFYHQTTEHKFQEQVSKELIGLVLTCKYNNKTYVDDIDWDQNP 243
 Db 270 SHKVLRSSETVLDPMFNFYHQTTEHKFQEQVSKELIGLVLTCKYNNKTYVDDIDWDQNP 329
 QY 244 SFFKADGSGSVSFLEYRKYNOEITDLKQPVLSQPKRRRGGTLPGPAMLIPELCYL 303
 Db 330 SFFKADGSGSVSFLEYRKYNOEITDLKQPVLSQPKRRRGGTLPGPAMLIPELCYL 389
 QY 304 TGLTDMKRNDFNVMKDLAVHTRLTPEQREVGELIDYIHKNNDVQRELDMGLSFDNL 363
 Db 390 TGLTDMKRNDFNVMKDLAVHTRLTPEQREVGELIDYIHKNNDVQRELDMGLSFDNL 449
 QY 364 LSFSGRILQTEKHOGGKTFDYNPOFADWSKETRGAPLISVKPLDNWLLIYTRNYYEAA 423
 Db 450 LSFSGRILQTEKHOGGKTFDYNPOFADWSKETRGAPLISVKPLDNWLLIYTRNYYEAA 509
 QY 424 SLIQNLFKVTPAMGMQMRKAIMIEVDRTTEAYLRVLQKVTADTQIVVCLLSNRKDYD 483
 Db 510 SLIQNLFKVTPAMGMQMRKAIMIEVDRTTEAYLRVLQKVTADTQIVVCLLSNRKDYD 569
 QY 484 AIKKYLCTDPTSPQCVARTLGKQQTWMAIATKIALQMNCKMGELWRVDIPLKLVMI 543
 Db 570 AIKKYLCTDPTSPQCVARTLGKQQTWMAIATKIALQMNCKMGELWRVDIPLKLVMI 629
 QY 544 GIDCVHDMTAGRSIAGFVASINEGTRWFSRCIFQDRGOELVDGLKVCLOALRAWNSC 603
 Db 630 GIDCVHDMTAGRSIAGFVASINEGTRWFSRCIFQDRGOELVDGLKVCLOALRAWNSC 689
 QY 604 NEYMPRSRIIYVRDGVGGQKLTIVNYEVPQFLDCLKSIIGRGNPRLTVIIVKRVNTRFF 663
 Db 690 NEYMPRSRIIYVRDGVGGQKLTIVNYEVPQFLDCLKSIIGRGNPRLTVIIVKRVNTRFF 749
 QY 664 AQSGRQLNPLPGTVIDVEVTRPEWYDFFIVSQAVRSVSGVSPHYNVIYDNSGLKPDHIQ 723
 Db 750 AQSGRQLNPLPGTVIDVEVTRPEWYDFFIVSQAVRSVSGVSPHYNVIYDNSGLKPDHIQ 809
 QY 724 RLTYKLCHIIYNNPVGIRVPAPCOYAHKLAFLVGSQSIHREPNSLSNRLYYL 775
 Db 810 RLTYKLCHIIYNNPVGIRVPAPCOYAHKLAFLVGSQSIHREPNSLSNRLYYL 861

RESULT 3

Q8NA60 PRELIMINARY; PRT; 829 AA.
 ID Q8NA60
 AC Q8NA60;
 DT 01-OCT-2002 (TrEMBLrel. 22, Created)
 DT 01-OCT-2002 (TrEMBLrel. 22, Last sequence update)
 DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
 DE Hypothetical protein FLJ35814 (piwi).
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 OC NCBI_TaxID=9606;
 RN [1]
 RP SEQUENCE FROM N.A.

```
RC TISSUE-Testis;
RA Oshima A., Takahashi-Fujii A., Tanase T., Imose N., Takeuchi K.,
RA Arita M., Musashino K., Yuuki H., Hara H., Sugiyama T., Irie R.,
RA Otsuki T., Sato H., Wakamatsu A., Ishii S., Yamamoto J., Isono Y.,
RA Kawai-Hio Y., Saito K., Nishikawa T., Kimura K., Yamashita H.,
RA Matsuo K., Nakamura Y., Sekine M., Kikuchi H., Kanda K., Wagatsuma M.,
RA Murakawa K., Kanehori K., Sugiyama A., Kawakami B., Suzuki Y.,
RA Sugano S., Nagahara K., Masuho Y., Nagai K., Isogai T.;
RT "NEDO human cDNA sequencing project.";
RL Submitted (JUL-2002) to the EMBL/GenBank/DBJ databases.
DR EMBL; AK093133; BAC04068.1; -.
DR InterPro; IPR003100; PAZ.
DR InterPro; IPR003165; Piwi.
DR Pfam; PF02170; PAZ; 1.
DR Pfam; PF02171; Piwi; 1.
DR PROSITE; PS50821; PAZ; 1.
DR PROSITE; PS50822; PIWI; 1.
KW Hypothetical protein.
KW SEQUENCE 829 AA; 94816 MW; 3AF9FD1D0FA9F5A7 CRC64;
Query Match 83.2%; Score 645; DB 4; Length 829;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 645; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 93 TEVSKTRNGEDVRITITLTNELPPTSPCLQFYNIIPRLKIMNLQOIGRNYNPNDP 152
Db |||||
QY 179 TEVSKTRNGEDVRITITLTNELPPTSPCLQFYNIIPRLKIMNLQOIGRNYNPNDP 238
Db |||||
QY 153 IDIPSHRLVWPGFTTSILQYENSIMLCTDVSHKVLRSSTVLDPMFNFYHQTBEHKFBQ 212
Db |||||
QY 239 IDIPSHRLVWPGFTTSILQYENSIMLCTDVSHKVLRSSTVLDPMFNFYHQTBEHKFBQ 298
Db |||||
QY 213 VSKELIGLVLTAKYNNKTVRVDIDWQNPSTFKKADGSEVSFLEYRKYQNOEITDLK 272
Db |||||
QY 299 VSKELIGLVLTAKYNNKTVRVDIDWQNPSTFKKADGSEVSFLEYRKYQNOEITDLK 358
Db |||||
QY 273 QPVLSQPKRRRGGTLPQAMLIPELCLVLTGTDKMDNFVMDKLVHTRLTPEQRQ 332
Db |||||
QY 359 QPVLSQPKRRRGGTLPQAMLIPELCLVLTGTDKMDNFVMDKLVHTRLTPEQRQ 418
Db |||||
QY 333 REVGRLLIDYHKNQNVQRELDMGLSFDNLSFSGRILQTEKHGGKTFDYNPOFADW 392
Db |||||
QY 419 REVGRLLIDYHKNQNVQRELDMGLSFDNLSFSGRILQTEKHGGKTFDYNPOFADW 478
Db |||||
QY 393 SKETRGAPLISVKPLDNWLLIYTRRNYEAANSILQNLFKVTPANGMQRKAIMIEVDRT 452
Db |||||
QY 479 SKETRGAPLISVKPLDNWLLIYTRRNYEAANSILQNLFKVTPANGMQRKAIMIEVDRT 538
Db |||||
QY 453 EAYLRVLOQKVATDTQIVVCLSSNRKDYDAIKKYLCTDCTPPSCVWARTLGKQOTVM 512
Db |||||
QY 539 EAYLRVLOQKVATDTQIVVCLSSNRKDYDAIKKYLCTDCTPPSCVWARTLGKQOTVM 598
Db |||||
QY 513 AIATKIALQNMCKMGELNRVDIPLKLVMIIVGIDCYHDMTAGRRSIAGFVASINEGTRW 572
Db |||||
QY 599 AIATKIALQNMCKMGELNRVDIPLKLVMIIVGIDCYHDMTAGRRSIAGFVASINEGTRW 658
Db |||||
QY 573 FSCIFQDRGQELVDGLKVCLOALRAWNSCNEIMPSRIIIVYRDGVDGQGLKTLVNYEVP 632
Db |||||
QY 659 FSCIFQDRGQELVDGLKVCLOALRAWNSCNEIMPSRIIIVYRDGVDGQGLKTLVNYEVP 718
Db |||||
QY 633 QFLDCLKSGRGNPRITVIVKRVNTRFFAQSGLRQNPGLPCTVIDVETPPEWYDFP 692
Db |||||
QY 719 QFLDCLKSGRGNPRITVIVKRVNTRFFAQSGLRQNPGLPCTVIDVETPPEWYDFP 778
Db |||||
QY 693 IVSQAVRSGSVSPHYNVIYDNSGLKPDHILQRLTYKLCHIIYINWP 737
Db |||||
QY 779 IVSQAVRSGSVSPHYNVIYDNSGLKPDHILQRLTYKLCHIIYINWP 823
Db |||||
RESULT 4
O95404 PRELIMINARY; PRT; 861 AA.
ID O95404
AC O95404;
```

```
QY 544 GIDCHDMTAGRSIAGFVASINEGTRWFSRCIFQDRGQELVDGLKYCLQALRAWNSC 603
DB 630 GIDCHDMTAGRSIAGFVASINEGTRWFSRCIFQDRGQELVDGLKYCLQALRAWNSC 689
QY 604 NEYMPESRIIVYRDGVDGOLKTLVNVYVPOFLDKLSIGRGYNPLTIVIVVKRWNTREFF 663
DB 690 NEYMPESRIIVYRDGVDGOLKTLVNVYVPOFLDKLSIGRGYNPLTIVIVVKRWNTREFF 749
QY 664 AOSGRLQNLPGTVIDVEVTRPEWYDFIVSQAVRSVSPSTHYNVYDNSGLKPDHIQ 723
DB 750 AOSGRLQNLPGTVIDVEVTRPEWYDFIVSQAVRSVSPSTHYNVYDNSGLKPDHIQ 809
QY 724 RLTYKLCCHIYNNWPGVIRVPAPCOYAHKLAFLVGQSIHREPNSLSNRLYYL 775
DB 810 RLTYKLCCHIYNNWPGVIRVPAPCOYAHKLAFLVGQSIHREPNSLSNRLYYL 861

RESULT 5
Q8TBYS PRELIMINARY; PRT; 861 AA.
AC Q8TBYS;
DT 01-JUN-2002 (TrEMBLrel. 21, Created)
DT 01-JUN-2002 (TrEMBLrel. 21, Last sequence update)
DT 01-OCT-2002 (TrEMBLrel. 22, Last annotation update)
DE PIWI-like 1 (Drosophila).
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]_
RP SEQUENCE FROM N.A.
RC TISSUE=Testis;
RA Strausberg R.;
RL Submitted (APR-2002) to the EMBL/GenBank/DBJ databases.
DB EMBL; BC028581; AA28581.1; -.
DR InterPro; IPR003100; PAZ.
DR InterPro; IPR003185; Piwi.
DR Pfam; PF02170; PAZ; 1.
DR Pfam; PF02171; Piwi; 1.
DR PROSITE; PS0821; PAZ; 1.
DR PROSITE; PS0822; PIWI; 1.
SQ SEQUENCE 861 AA; 98545 MW; D33376EDED743A CRC64;

Query Match 60.5%; Score 469; DB 4; Length 861;
Best Local Similarity 99.6%; Pred. No. 0;
Matches 769; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 4 GVNTQNLDHVKESKTGSSGIIIVRLSTNHFLTSRPOWALYQVHIDYNPMEARRLSAL 63
DB 90 GVNTQNLDHVKESKTGSSGIIIVRLSTNHFLTSRPOWALYQVHIDYNPMEARRLSAL 149
QY 64 LFQHEDLIGKCHAFDGTILFLPKELQKQVTEVFSKTRNGEDVRAITITLNLPTPTCL 123
DB 150 LFQHEDLIGKCHAFDGTILFLPKELQKQVTEVFSKTRNGEDVRAITITLNLPTPTCL 209
QY 124 QFYNIIFRLLKMNLOQIGRNYNPNPDIDISHRLVWPGTTSILOYSIMLCTDV 183
DB 210 QFYNIIFRLLKMNLOQIGRNYNPNPDIDISHRLVWPGTTSILOYSIMLCTDV 269
QY 184 SHKVLRSSETVLDNFNFYHOTEHKEFQVSKELIGLWLTCKNNKTYRVDIDWDQNPX 243
DB 270 SHKVLRSSETVLDNFNFYHOTEHKEFQVSKELIGLWLTCKNNKTYRVDIDWDQNPX 329
QY 244 STFKKADGSEVSFLYYRKQYNQIBITLQKPLVVSQPKRRRGPGGTLPGPAMLIPELCYL 303
DB 330 STFKKADGSEVSFLYYRKQYNQIBITLQKPLVVSQPKRRRGPGGTLPGPAMLIPELCYL 389
QY 304 TGLTDKMRNDFNVKDLAVHTRLTPEQORQVGLIDYIHKQNVQRELWDGWSLPSNL 363
DB 390 TGLTDKMRNDFNVKDLAVHTRLTPEQORQVGLIDYIHKQNVQRELWDGWSLPSNL 449
QY 364 LSPSGRILQTEKIHOGGKTFDYNPQFADWSKETRGAPLISVKNPLDNMLLIYTRRNYEAA 423
```

```
DB 450 LSPSGRILQTEKIHOGGKTFDYNPQFADWSKETRGAPLISVKNPLDNMLLIYTRRNYEAA 509
QY 424 SLIQNLKFKVTPAMQWQKKAIMEIVDDRTETAYLVLQOKVTADTQIVVCLLSNRKDKYD 483
DB 510 SLIQNLKFKVTPAMQWQKKAIMEIVDDRTETAYLVLQOKVTADTQIVVCLLSNRKDKYD 569
QY 484 AIKKYLCTDPTSCQCVVARTLGHQQTVMATATKIALQMNCKMGELWRVDIPLKLVIV 543
DB 570 AIKKYPTDCTPSCQCVVARTLGHQQTVMATATKIALQMNCKMGELWRVDIPLKLVIV 629
QY 544 GIDCHDMTAGRSIAGFVASINEGTRWFSRCIFQDRGQELVDGLKYCLQALRAWNSC 603
DB 630 GIDCHDMTAGRSIAGFVASINEGTRWFSRCIFQDRGQELVDGLKYCLQALRAWNSC 689
QY 604 NEYMPESRIIVYRDGVDGOLKTLVNVYVPOFLDKLSIGRGYNPLTIVIVVKRWNTREFF 663
DB 690 NEYMPESRIIVYRDGVDGOLKTLVNVYVPOFLDKLSIGRGYNPLTIVIVVKRWNTREFF 749
QY 664 AOSGRLQNLPGTVIDVEVTRPEWYDFIVSQAVRSVSPSTHYNVYDNSGLKPDHIQ 723
DB 750 AOSGRLQNLPGTVIDVEVTRPEWYDFIVSQAVRSVSPSTHYNVYDNSGLKPDHIQ 809
QY 724 RLTYKLCCHIYNNWPGVIRVPAPCOYAHKLAFLVGQSIHREPNSLSNRLYYL 775
DB 810 RLTYKLCCHIYNNWPGVIRVPAPCOYAHKLAFLVGQSIHREPNSLSNRLYYL 861

RESULT 6
Q9JWB7 PRELIMINARY; PRT; 862 AA.
AC Q9JWB7;
DT 01-OCT-2000 (TrEMBLrel. 15, Created)
DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DT 01-OCT-2002 (TrEMBLrel. 22, Last annotation update)
DE MIWI (piwi).
GN PIWI1 OR MIWI.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]_
RP SEQUENCE FROM N.A.
RA Miyagawa S.K., Kimura T., Nakano T.;
RT "Molecular Cloning and characterization of piwi family genes.";
RL Submitted (SEP-1999) to the EMBL/GenBank/DBJ databases.
RN [2]_
RP SEQUENCE FROM N.A.
RC TISSUE=Testis;
RA Deng W., Lin H.;
RT "Miwi, a murine homolog of piwi, encodes a cytoplasmic protein
essential for spermatogenesis.";
RL Submitted (OCT-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; AB032604; BAA93705.1; -.
DR EMBL; AF438405; AAL31014.1; -.
DR MGD; MGI:1928897; Piwi1.
DR InterPro; IPR003100; PAZ.
DR InterPro; IPR003185; Piwi.
DR Pfam; PF02170; PAZ; 1.
DR Pfam; PF02171; Piwi; 1.
DR PROSITE; PS0821; PAZ; 1.
DR PROSITE; PS0822; PIWI; 1.
SQ SEQUENCE 862 AA; 98574 MW; 45588D13284CCC4C CRC64;

Query Match 15.9%; Score 123; DB 11; Length 862;
Best Local Similarity 100.0%; Pred. No. 9.3e-120;
Matches 123; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 222 VLTKNKTYRVDIDWDQNPXSTFKKADGSEVSFLYYRKQYNQIBITLQKPLVVSQPK 281
DB 309 VLTKNKTYRVDIDWDQNPXSTFKKADGSEVSFLYYRKQYNQIBITLQKPLVVSQPK 368
QY 282 RRRPGGTLPGPAMLIPELCYLTGLTDKMRNDFNVKDLAVHTRLTPEQORQVGLIDY 341
```

```

Db 369 RRRGGGTLPGFAMLIPELCYLTLGLTKQNRDNENWMDLAHVHTRLPQQRQREVGRLIDY 428
Qy 342 IHK 344
Db 429 IHK 431

RESULT 7
Q9GPA7 ID Q8UVX0 PRELIMINARY; PRT; 858 AA.
AC Q8UVX0;
DT 01-MAR-2002 (TREMELrel. 20, Created)
DT 01-MAR-2002 (TREMELrel. 20, Last sequence update)
DT 01-OCT-2002 (TREMELrel. 22, Last annotation update)
DE Piwi protein.
OS Brachydanio rerio (zebrafish) (Danio rerio).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Actinopterygii; Neopterygii; Teleostei; Ostariophysi; Cypriniformes;
OC Cyprinidae; Danio.
OX NCBI_TaxID=7955;
RN [1]
RP SEQUENCE FROM N.A.
RA Weeratne S.D., Gong Z., Tan C.-H.;
RT "Cloning and characterization of zebrafish homolog of piwi, essential
for germ-line stem cell self-renewal.";
RL Submitted (JAN-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF36369; AAL57170.1; -.
DR InterPro; IPR003100; PAZ.
DR InterPro; IPR003165; Piwi.
DR Pfam; PF02170; PAZ; 1.
DR Pfam; PF02171; Piwi; 1.
DR PROSITE; PS50821; PAZ; 1.
DR PROSITE; PS50822; PIWI; 1.
SQ SEQUENCE 858 AA; 97451 MW; 6A12F2E511465777 CRC64;

Query Match 2.7%; Score 21; DB 13; Length 858;
Best Local Similarity 100.0%; Pred. No. 1.6e-12;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 741 RVPAPCOYAHKLAFLVQSQIH 761
Db 824 RVPAPCOYAHKLAFLVQSQIH 844

RESULT 8
Q9GPA7 ID Q9GPA7 PRELIMINARY; PRT; 808 AA.
AC Q9GPA7;
DT 01-MAR-2001 (TREMELrel. 16, Created)
DT 01-MAR-2001 (TREMELrel. 16, Last sequence update)
DT 01-OCT-2002 (TREMELrel. 22, Last annotation update)
DE Seawi (Fragment).
OS Strongylocentrotus purpuratus (Purple sea urchin).
OC Eukaryota; Metazoa; Echinodermata; Eleutherozoa; Echinozoa;
OC Echinoidea; Euechinoidea; Echinacea; Echinoida; Strongylocentrotidae;
OC Strongylocentrotus.
OX NCBI_TaxID=7668;
RN [1]
RP SEQUENCE FROM N.A.
RA Rodriguez A.J., Bonder E.M.;
RT "Seawi - Cloning and Molecular Characterization of a Sea Urchin
Homolog of Piwi.";
RL Submitted (NOV-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL; AY014900; AAG42534.1; -.
DR InterPro; IPR003100; PAZ.
DR InterPro; IPR003165; Piwi.
DR Pfam; PF02170; PAZ; 1.
DR Pfam; PF02171; Piwi; 1.
DR PROSITE; PS50821; PAZ; 1.
DR PROSITE; PS50822; PIWI; 1.
NON TER 808
SQ SEQUENCE 808 AA; 91366 MW; 19528F4B9D10474C CRC64;

```

```

Query Match 2.1%; Score 16; DB 5; Length 808;
Best Local Similarity 100.0%; Pred. No. 2.7e-07;
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 222 VLTKNKNTYRVDDID 237
Db 302 VLTKNKNTYRVDDID 317

RESULT 9
Q9GPA8 ID Q9GPA8 PRELIMINARY; PRT; 854 AA.
AC Q9GPA8;
DT 01-MAR-2001 (TREMELrel. 16, Created)
DT 01-MAR-2001 (TREMELrel. 16, Last sequence update)
DT 01-OCT-2002 (TREMELrel. 22, Last annotation update)
DE Seawi.
OS Strongylocentrotus purpuratus (Purple sea urchin).
OC Eukaryota; Metazoa; Echinodermata; Eleutherozoa; Echinozoa;
OC Echinoidea; Euechinoidea; Echinacea; Echinoida; Strongylocentrotidae;
OC Strongylocentrotus.
OX NCBI_TaxID=7668;
RN [1]
RP SEQUENCE FROM N.A.
RA Rodriguez A.J., Bonder E.M.;
RT "Seawi - Cloning and Molecular Characterization of a Sea Urchin
Homolog of Piwi.";
RL Submitted (NOV-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL; AY014899; AAG42533.1; -.
DR InterPro; IPR003100; PAZ.
DR InterPro; IPR003165; Piwi.
DR Pfam; PF02170; PAZ; 1.
DR Pfam; PF02171; Piwi; 1.
DR PROSITE; PS50821; PAZ; 1.
DR PROSITE; PS50822; PIWI; 1.
SQ SEQUENCE 854 AA; 96722 MW; 509A1D39C0D1922C CRC64;

Query Match 2.1%; Score 16; DB 5; Length 854;
Best Local Similarity 100.0%; Pred. No. 2.9e-07;
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 222 VLTKNKNTYRVDDID 237
Db 302 VLTKNKNTYRVDDID 317

RESULT 10
Q17567 ID Q17567 PRELIMINARY; PRT; 722 AA.
AC Q17567;
DT 01-NOV-1996 (TREMELrel. 01, Created)
DT 01-NOV-1996 (TREMELrel. 01, Last sequence update)
DT 01-MAR-2003 (TREMELrel. 23, Last annotation update)
DE C01G5.2 protein.
GN C01G5.2.
OS Caenorhabditis elegans.
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;
OC Rhabditidae; Peloderinae; Caenorhabditis.
OX NCBI_TaxID=6239;
RN [1]
RP SEQUENCE FROM N.A.
RA STRAIN=Bristol N2;
RX MEDLINE=99069613; PubMed=9851916;
RA None;
RT "Genome sequence of the nematode C. elegans: a platform for
investigating biology. The C. elegans Sequencing Consortium.";
RL Science 282:2012-2018(1998).
RN [2]
RP SEQUENCE FROM N.A.
RA STRAIN=Bristol N2;
RA Bradshaw H., Stellyes L.;
RT "The sequence of C. elegans cosmid C01G5.";
RL Submitted (MAR-1996) to the EMBL/GenBank/DBJ databases.

```

```

Query Match      1.4%; Score 11; DB 4; Length 666;
Best Local Similarity 100.0%; Pred. NO. 0.042;
Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy      742 VPAPCQYAHKL 752
      |||||
Db      633 VPAPCQYAHKL 643

RESULT 14
Q8N9V8

```

```

ID Q8N9V8 PRELIMINARY; PRT; 852 AA.
AC Q8N9V8;
DT 01-OCT-2002 (TrEMBLrel. 22, Created)
DT 01-OCT-2002 (TrEMBLrel. 22, Last sequence update)
DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE Hypothetical protein FLJ36156 (piwi).
OS Homo sapiens (Human)
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Testis;
RA Ishibashi T., Kanehori K., Yosida M., Watanabe S., Ishida S., Ono Y.,
RA Hotta T., Hiraoka S., Murakawa K., Takiguchi S., Kusano J., Chiba Y.,
RA Watanabe M., Fujimori K., Tanai H., Ishida M., Yamashita H., Chiba Y.,
RA Sugiyama T., Irie R., Otsuki T., Sato H., Wakamatsu A., Ishii S.,
RA Yamamoto J., Isono Y., Kawai-Hio Y., Saito K., Nishikawa T.,
RA Kimura K., Matsuo K., Nakamura Y., Sekine M., Kikuchi H., Kanda K.,
RA Wagatsuma M., Takahashi-Fujii A., Oshina A., Sugiyama A., Kawakami B.,
RA Suzuki Y., Sugano S., Nagahari K., Masuno Y., Negai K., Isogai T.;
RA Submitted (JUL-2002) to the EMBL/GenBank/DBJ databases.
RT "NEDO human cDNA sequencing project."
RL EMBL; AK093475; BAC04179.1; -
DR InterPro; IPR003100; PAZ
DR InterPro; IPR003185; Piwi.
DR Pfam; PF02170; PAZ; 1.
DR Pfam; PF02171; Piwi; 1.
DR PROSITE; PS0821; PAZ; 1.
DR PROSITE; PS0822; PIWI; 1.
KW Hypothetical protein.
SQ SEQUENCE 852 AA; 96560 MW; 75C6FEFAE70701B CRC64;

Query Match 1.4%; Score 11; DB 4; Length 852;
Best Local Similarity 100.0%; Pred.No. 0.052;
Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 742 VPAPCOYAHKL 752
DB 819 VPAPCOYAHKL 829

RESULT 15
Q8NEH2 PRELIMINARY; PRT; 852 AA.
AC Q8NEH2;
DT 01-OCT-2002 (TrEMBLrel. 22, Created)
DT 01-OCT-2002 (TrEMBLrel. 22, Last sequence update)
DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE Similar to piwi like homolog 1 (Drosophila).
OS Homo sapiens (Human)
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Testis;
RA Strausberg R.;
RL Submitted (JUN-2002) to the EMBL/GenBank/DBJ databases.
DR EMBL; BC031060; AAH31060.1; -
DR InterPro; IPR003100; PAZ.
DR InterPro; IPR003165; Piwi.
DR Pfam; PF02170; PAZ; 1.
DR Pfam; PF02171; Piwi; 1.
DR PROSITE; PS0821; PAZ; 1.
DR PROSITE; PS0822; PIWI; 1.
SQ SEQUENCE 852 AA; 96588 MW; 37769EE078B96D13 CRC64;

Query Match 1.4%; Score 11; DB 4; Length 852;
Best Local Similarity 100.0%; Pred.No. 0.052;
Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 742 VPAPCOYAHKL 752
DB 819 VPAPCOYAHKL 829
```

DB 819 VPAPCOYAHKL 829

Search completed: December 10, 2003, 17:33:27
Job time : 46 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: December 10, 2003, 17:32:26 ; Search time 18 Seconds

(without alignments)
2024.761 Million cell updates/sec

Title: US-10-043-774B-2

Perfect score: 775

Sequence: 1 MIFGVNTRQNLHVKESTG.....VQGSIHREPNLSNRLYYL 775

Scoring table:

Gapop 60.0 , Gapext 60.0

Searched: 127863 seqs, 47026705 residues

Word size : 0

Total number of hits satisfying chosen parameters: 127863

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database : SwissProt_41.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	9	1.2	1040	1 YO43 CAEEL	P34681 caenorhabdi
2	8	1.0	320	1 Y381 METJA	Q57826 methanococ
3	8	1.0	369	1 WNT1 ANEME	P21551 ambystoma m
4	8	1.0	520	1 ECHE HELMO	Q29258 heliobacill
5	7	0.9	112	1 YN48 ARCFC	O30321 archaeoglob
6	7	0.9	130	1 YJID ECOLI	Q91029 gallus gall
7	7	0.9	134	1 WNT1 CHICK	P39375 escherichia
8	7	0.9	135	1 YJ01 AQUAE	O67739 aquifex aeo
9	7	0.9	148	1 VSN1 NEIMC	Q91rm5 neisseria m
10	7	0.9	149	1 VB15 VACCC	P21089 vaccinia vi
11	7	0.9	149	1 VB15 VACCV	P24772 vaccinia vi
12	7	0.9	149	1 VB15 VARV	P33877 variola vir
13	7	0.9	207	1 YFAT ECOLI	P76466 escherichia
14	7	0.9	209	1 ABPA PRUPE	Q92ra4 prunus pers
15	7	0.9	209	1 ABGP PRUPE	O04012 prunus pers
16	7	0.9	209	1 ENG1 NEIGO	O05132 neisseria g
17	7	0.9	244	1 YBGL ECOLI	P75746 escherichia
18	7	0.9	274	1 DAPF HABIN	P44859 haemophilus
19	7	0.9	274	1 DAPF PASMU	P57962 pasteurella
20	7	0.9	274	1 DAPF SALTY	Q91696 salmonella
21	7	0.9	274	1 DAPF YERPE	P46357 yersinia pe
22	7	0.9	275	1 CHER VIBAN	O57508 vibrio angu
23	7	0.9	275	1 CHER VIBCH	Q5x9k2 vibrio para
24	7	0.9	275	1 CHRI1 VIBCH	Q9Kq06 vibrio chol
25	7	0.9	284	1 TLX2 HUMAN	O43763 homo sapien
26	7	0.9	284	1 TLX2 MOUSE	O61663 mus musculu
27	7	0.9	284	1 Y766 AQUAE	O65965 aquifex aeo
28	7	0.9	288	1 YQD5 CAEEL	Q09285 caenorhabdi
29	7	0.9	291	1 TLX3 HUMAN	O43711 homo sapien
30	7	0.9	291	1 TLX3 MOUSE	O55144 mus musculu
31	7	0.9	294	1 SC17 PICPA	Q9p4d0 pichia past
32	7	0.9	294	1 YJ05 YEAST	P47008 saccharomyc
33	7	0.9	297	1 TLX1 CHICK	O93366 gallus gall

34	7	0.9	297	1 TLX3 CHICK	O93367 gallus gall
35	7	0.9	319	1 PFSN ECOLI	P29131 escherichia
36	7	0.9	329	1 RRP4 SCHPO	Q09704 schizosacch
37	7	0.9	330	1 TLX1 HUMAN	P31314 homo sapien
38	7	0.9	332	1 TLX1 MOUSE	P43345 mus musculu
39	7	0.9	335	1 AROB PYRFU	Q8u0a8 pyrococcus
40	7	0.9	338	1 AROB PYRAB	Q9v1h9 pyrococcus
41	7	0.9	356	1 VF7 ERD	P35935 broadhaven
42	7	0.9	357	1 CAD2 PICAB	O82035 picea abies
43	7	0.9	357	1 CAD7 PICAB	Q08350 picea abies
44	7	0.9	357	1 CADH PINRA	Q09076 pinus radia
45	7	0.9	357	1 CADH PINTA	P41637 pinus taeda

ALIGNMENTS

RESULT 1
YO43 CAEEL
ID YO43 CAEEL STANDARD; PRT; 1040 AA.
AC P34681;
DT 01-FEB-1994 (Rel. 28, Created)
DT 01-FEB-1995 (Rel. 31, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Hypothetical protein ZK757.3 in chromosome III.
GN ZK757.3
OS Caenorhabditis elegans.
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;
OC Rhabditidae; Peloderinae; Caenorhabditis.
OX NCBI TaxID=6239;
[1] _SEQUENCE FROM N.A.
RN STRAIN=Bristol N2;
RC MEDLINE=94150718; PubMed=7906398;
RA Wilson R., Almscough R., Anderson K., Baynes C., Berks M., Coulson A.,
RA Bonfield J., Burton J., Connell M., Copsey T., Cooper J., Fraser A.,
RA Craxton M., Dear S., Du Z., Durbin R., Favello A., Fraser A.,
RA Fulton L., Gardner A., Green P., Hawkins T., Hillier L., Jier M.,
RA Johnston L., Jones M., Kershaw J., Kirsten J., Laister N.,
RA Latreille P., Lightning J., Lloyd C., Mortimore B., O'Callaghan M.,
RA Parsons J., Percy C., Rifkin L., Roopra A., Saunders D., Showken R.,
RA Sims M., Smalton N., Smith A., Smith M., Sonhammer E., Staden R.,
RA Sulston J., Thierry-Mieg J., Thomas K., Vaudin M., Vaughan K.,
RA Waterston R., Watson A., Weinstock L., Wilkinson-Sproat J.,
RA Wohldman P., Watson A., Weinstock L., Wilkinson-Sproat J.,
RT "2.2 Mb of contiguous nucleotide sequence from chromosome III of C.
RT elegans.";
RL Nature 368:32-38(1994).
CC -!- SIMILARITY: Contains 1 PAZ domain.
CC -!- SIMILARITY: Contains 1 Piwi domain.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; Z30215; CAA82941.1; -
CC EMBL; Z29121; CAA82389.1; JOINED.
CC EMBL; Z29121; CAA82389.1; -
CC EMBL; Z30215; CAA82389.1; JOINED.
CC PIR; D88568; D88568.
CC WormPep; ZK757.3; CE01117.
CC InterPro; IPR003100; PAZ.
CC InterPro; IPR003165; Piwi.
CC Pfam; PF02170; PAZ; 1.
CC Pfam; PF02171; Piwi; 1.
CC PROSITE; P850821; PAZ; 1.
CC PROSITE; P850822; PIWI; 1.
KW Hypothetical protein.
FT DOMAIN 378 486 PAZ.

```

FT DOMAIN      660 966      PIWI.
SQ SEQUENCE    1040 AA; 115415 MW; 4C8483C1F1D72338 CRC64;

Query Match      1.2%; Score 9; DB 1; Length 1040;
Best Local Similarity 100.0%; Pred. No. 1;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      610 RIIVRDGV 618
DB      811 RIIVRDGV 819

RESULT 2
Y381_METJUA      STANDARD;      PRT;      320 AA.
AC      Q57826;
DT      01-NOV-1997 (Rel. 35, Created)
DT      01-NOV-1997 (Rel. 35, Last sequence update)
DT      16-OCT-2001 (Rel. 40, Last annotation update)
DE      Hypothetical protein M30381.
GN      M30381.
OS      Methanococcus jannaschii.
OC      Archaea; Euryarchaeota; Methanococci; Methanococcales;
OC      Methanocaldococcaceae; Methanocaldococcus.
OX      NCBI_TaxID=2190;
RN      [1]
SEQUENCE FROM N.A.
RC      STRAIN=JAL-1 / DSM 2661 / ATCC 43067;
RX      MEDLINE=96337999; PubMed=8688087;
RA      Bult C.J., White O., Olsen G.J., Zhou L., Fleischmann R.D.,
RA      Sutton G.G., Blake J.A., Fitzgerald L.M., Clayton R.A., Gocayne J.D.,
RA      Karlavage A.R., Dougherty B.A., Tomb J.F., Adams M.D., Reich C.I.,
RA      Overbeek R., Kirkness E.F., Weinstock K.G., Merrick J.M., Glodek A.,
RA      Scott J.L., Geoghagen N.S.M., Weidman J.F., Fuhrmann J.L., Nguyen D.,
RA      Uterback T.R., Kelley J.M., Peterson J.D., Sadow P.W., Hanna M.C.,
RA      Cotton M.D., Roberts K.M., Hurst M.A., Kaine B.P., Borodovsky M.,
RA      Klenk H.-P., Fraser C.M., Smith H.O., Woese C.R., Venter J.C.;
RT      "Complete genome sequence of the methanogenic archaeon, Methanococcus
RT      jannaschii.";
RL      Science 273:1058-1073 (1996).

CC      This SWISS-PROT entry is copyright. It is produced through a collaboration
CC      between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC      the European Bioinformatics Institute. There are no restrictions on its
CC      use by non-profit institutions as long as its content is in no way
CC      modified and this statement is not removed. Usage by and for commercial
CC      entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC      or send an email to license@isb-sib.ch).
CC      -----
CC      EMBL; U67491; AAB98370.1; -.
CC      PIR; E64347; E64347.
CC      TIGR; M30381; -.
CC      InterPro; IPR002764; DUF73.
CC      Pfam; PF01905; DUF73; 1.
CC      ProDom; PD017873; DUF73; 1.
CC      Hypothetical protein; Complete proteome.
SQ      SEQUENCE 320 AA; 35924 MW; 81C8DE178F4B223 CRC64;

Query Match      1.0%; Score 8; DB 1; Length 320;
Best Local Similarity 100.0%; Pred. No. 4.2;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      247 KKADGSEV 254
DB      95 KKADGSEV 102

RESULT 3
WNT1_AMBME
ID      WNT1_AMBME      STANDARD;      PRT;      369 AA.
AC      P21551;
DT      01-MAY-1991 (Rel. 18, Created)
DT      01-MAY-1991 (Rel. 18, Last sequence update)
DT

```

```

DT      15-SEP-2003 (Rel. 42, Last annotation update)
DE      Wnt-1 protein precursor.
GN      WNT-1.
OS      Ambystoma mexicanum (Axolotl).
OC      Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC      Amphibia; Batrachia; Caudata; Salamandroidae; Ambystomatidae;
OC      Ambystoma.
OX      NCBI_TaxID=8296;
RN      [1]
SEQUENCE FROM N.A.
RC      MEDLINE=91081334; PubMed=2259633;
RA      Busse U., Guay J., Seguin C.;
RT      "Nucleotide sequence of a cDNA encoding Wnt-1 of the Mexican axolotl
RT      Ambystoma mexicanum.";
RL      Nucleic Acids Res. 18:7439-7439 (1990).
RN      [2]
ERRATUM.
RC      MEDLINE=91204483; PubMed=2017393;
RA      Busse U., Guay J., Seguin C.;
RL      Nucleic Acids Res. 19:981-981 (1991).
RN      [3]
CHARACTERIZATION.
RC      MEDLINE=93285407; PubMed=8508949;
RA      Busse U., Seguin C.;
RT      "Molecular analysis of the Wnt-1 proto-oncogene in Ambystoma
RT      mexicanum (axolotl) embryos.";
RL      Differentiation 53:7-15 (1993).
CC      -!- FUNCTION: LIGAND FOR MEMBERS OF THE FRIZZLED FAMILY OF SEVEN
CC      TRANSMEMBRANE RECEPTORS. PROBABLE DEVELOPMENTAL PROTEIN. MAY BE A
CC      SIGNALING MOLECULE IMPORTANT IN CNS DEVELOPMENT. IS LIKELY TO
CC      SIGNAL OVER ONLY FEW CELL DIAMETERS.
CC      -!- SUBCELLULAR LOCATION: Possibly secreted and associates with the
CC      extracellular matrix.
CC      -!- DEVELOPMENTAL STAGE: EARLY BLASTULA UNTIL GASTRULATION, BARELY
CC      EXPRESSED DURING GASTRULATION AND PRESENT AGAIN FROM NEURULATION
CC      UNTIL LATE EMBRYOGENESIS.
CC      -!- SIMILARITY: Belongs to the Wnt family.
CC      -----
CC      This SWISS-PROT entry is copyright. It is produced through a collaboration
CC      between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC      the European Bioinformatics Institute. There are no restrictions on its
CC      use by non-profit institutions as long as its content is in no way
CC      modified and this statement is not removed. Usage by and for commercial
CC      entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC      or send an email to license@isb-sib.ch).
CC      -----
CC      EMBL; X55270; CAA38991.1; -.
CC      PIR; S13721; S13721.
CC      InterPro; IPR005817; Wnt.
CC      InterPro; IPR005816; Wnt_grthfactor.
CC      Pfam; PF00110; wnt; 1.
CC      PRINTS; PR01349; WNTPROTEIN.
CC      SMART; SM00097; WNT1; 1.
CC      PROSITE; PS00246; WNT1; 1.
CC      Wnt signaling pathway; Developmental protein; Glycoprotein; Signal.
FT      SIGNAL      1 19
FT      CHAIN       20 369
FT      CARBOHYD    28 278
FT      CARBOHYD    277 277
FT      CARBOHYD    315 315
FT      CARBOHYD    358 358
FT      SEQUENCE   369 AA; 41383 MW; DC215A620F619321 CRC64;
SQ

Query Match      1.0%; Score 8; DB 1; Length 369;
Best Local Similarity 100.0%; Pred. No. 4.7;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      282 RRGPGGT 289
DB      155 RRGPGGT 162

RESULT 4

```

```

BCHB_HELMO
ID BCHB_HELMO STANDARD; PRT; 520 AA.
AC Q9ZGE8;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Light-independent protochlorophyllide reductase subunit B
DE (EC 1.18.-.-) (LI-POR subunit B) (DPOR subunit B).
GN BCHB.

OS Helicobacillus mobilis.
OC Bacteria; Firmicutes; Clostridia; Clostridiales; Helicobacteriaceae;
OC Helicobacillus.
OX NCBI_TaxID=28064;

RN [1]
SEQUENCE FROM N.A.
RX MEDLINE=9061957; PubMed=9843979;
RA Xiong J., Inoue K., Bauer C.E.;
RT "Tracking molecular evolution of photosynthesis by characterization of
a major photosynthesis gene cluster from Helicobacillus mobilis.";
RL Proc. Natl. Acad. Sci. U.S.A. 95:14851-14856(1998).
CC protochlorophyllide (pchlide) to form chlorophyllide a (Chlide)
CC (By similarity). This reaction is light-independent.
CC -!- PATHWAY: Light-independent bacteriochlorophyll biosynthesis.
CC -!- SUBUNIT: Protobacteriochlorophyllide reductase is thought to be composed
CC of three subunits; bchL, bchN and bchB. Could form a
CC heterotrimer of two bchN and two bchN subunits.
CC -!- SIMILARITY: Belongs to the chlB / bchB / bchZ family.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).

EMBL; AF080002; AAC84030.1; ALT_INIT.
DR HAMAP; MF_00353; -.
DR InterPro; IPR000510; Oxred nitrogsnel.
DR InterPro; IPR005969; Protocchl reductB.
DR Pfam; PF00148; Oxigred nitro; 1.
DR TIGRFAMs; TIGR01278; DPOR BchB; 1.
DR Oxidoreductase, Photosynthesis; Bacteriochlorophyll biosynthesis.
SQ SEQUENCE 520 AA; 57775 MW; 7A2AC5FAEC8A4BD CRC64;

Query Match 1.0%; Score 8; DB 1; Length 520;
Best Local Similarity 100.0%; Pred.No. 6.4;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 501 VARTLGKQ 508
DB 107 VARTLGKQ 114

RESULT 5
YN48_ARCFU STANDARD; PRT; 112 AA.
AC O30321;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Hypothetical protein AF2348.
GN AF2348.
OS Archaeoglobus fulgidus.
OC Archaea; Euryarchaeota; Archaeoglobi; Archaeoglobales;
OC Archaeoglobaceae; Archaeoglobus.
OX NCBI_TaxID=2234;
RN [1]
SEQUENCE FROM N.A.
RX STRAIN=VC-16 / DSM 4304 / ATCC 49558;
RX MEDLINE=98049343; PubMed=9389475;
RA Klenk H.-P., Clayton R.A., Tomb J.-F., White O., Nelson K.E.,

```

```

RA Ketchum K.A., Dodson R.J., Gwinn M., Hickey E.K., Peterson J.D.,
RA Richardson D.L., Kervavage A.R., Graham D.E., Kyripides N.C.,
RA Fleischmann R.D., Quackenbush J., Lee N.H., Sutton G.G., Gill S.,
RA Kirkness E.F., Dougherty B.A., McKenney K., Adams M.D., Loftus B.,
RA Peterson S., Reich C.I., McNeil L.K., Badger J.H., Glöckle A., Zhou L.,
RA Overbeek R., Gocayne J.D., Weidman J.F., McDonald L., Utterback T.,
RA Cotton M.D., Spriggs T., Artiach P., Kaine B.P., Sykes S.M.,
RA Sadow P.W., D'Andrea K.P., Bowman C., Fujii C., Garland S.A.,
RA Mason T.M., Olsen G.J., Fraser C.M., Smith H.O., Woese C.R.,
RA Venter J.C.;
RT "The complete genome sequence of the hyperthermophilic, sulphate-
RT reducing archaeon Archaeoglobus fulgidus.";
RL Nature 390:364-370(1997).

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).

EMBL; AB001114; AAB91315.1; -.
DR PIR; D69543; D69543.
DR TIGR; AF2348; -.
KW Hypothetical protein; Complete proteome.
SQ SEQUENCE 112 AA; 13012 MW; C9C6AC0ACD6ACT730 CRC64;

Query Match 0.9%; Score 7; DB 1; Length 112;
Best Local Similarity 100.0%; Pred.No. 19;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 338 LIDYIHK 344
DB 34 LIDYIHK 40

RESULT 6
YJID_ECOLI STANDARD; PRT; 130 AA.
ID YJID_ECOLI
AC P39375;
DT 01-FEB-1995 (Rel. 31, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Hypothetical protein yjID.
GN yjID OR B4326.
OS Escherichia coli.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC Enterobacteriaceae; Escherichia.
OX NCBI_TaxID=562;
RN [1]
SEQUENCE FROM N.A.
RX STRAIN=KL12 / MG1655;
RX MEDLINE=95334362; PubMed=7610040;
RA Burland V.D., Plunkett G. III, Sofia H.J., Daniels D.L.,
RA Blattner F.R.;
RT "Analysis of the Escherichia coli genome VI: DNA sequence of the
RT region from 92.8 through 100 minutes.";
RL Nucleic Acids Res. 23:2105-2119(1995).

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).

EMBL; U14003; AAA97222.1; ALT_INIT.
DR EMBL; AE000503; AAC77282.1; ALT_INIT.
DR EcoGene; EG12565; yjID.
KW Hypothetical protein; Complete proteome.
SQ SEQUENCE 130 AA; 14747 MW; AA07645C39525D90 CRC64;

```

Query Match 0.9%; Score 7; DB 1; Length 130;
Best Local Similarity 100.0%; Pred. No. 21;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 59 LRSALLF 65
DB 69 LRSALLF 75

RESULT 7

WNT1_CHICK STANDARD; PRT; 134 AA.
ID WNT1_CHICK AC Q91029;
DT 15-JUL-1998 (Rel. 36, Created)
DT 15-JUL-1998 (Rel. 36, Last sequence update)
DT 15-SEP-2003 (Rel. 42, Last annotation update)
DE Wnt-1 protein (Fragment).
GN WNT-1.
OS Gallus gallus (Chicken).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Archosauria; Aves; Neognathae; Galliformes; Phasianinae;
OC Gallus.
OX NCBI_TaxID=9031;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=White leghorn; TISSUE=Embryo;
RX MEDLINE=95121198; PubMed=7821210;
RA Bally-Cuif L., Wassef M.;
RT "Ectopic induction and reorganization of Wnt-1 expression in quail/chick chimeras".
RL Development 120:3379-3394 (1994).

CC -!- FUNCTION: LIGAND FOR MEMBERS OF THE FRIZZLED FAMILY OF SEVEN TRANSMEMBRANE RECEPTORS. PROBABLE DEVELOPMENTAL PROTEIN. MAY BE A SIGNALING MOLECULE IMPORTANT IN CNS DEVELOPMENT. IS LIKELY TO SIGNAL OVER ONLY FEW CELL DIAMETERS. PROMINENT ROLE IN THE INDUCTION OF THE MESENCEPHALON AND CEREBELLUM.
CC -!- SUBCELLULAR LOCATION: Possibly secreted and associates with the extracellular matrix.
CC -!- DEVELOPMENTAL STAGE: EXPRESSION IN THE MET-MESENCEPHALIC REGION.
CC -!- SIMILARITY: Belongs to the Wnt family.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).
CC -----

DR EMBL; X81693; CAA57341.1; -.
DR F0150729; I50729.
DR InterPro; IPR005817; Wnt.
DR IntraPro; IPR005816; wnt_gthfactor.
DR Pfam; PF00110; wnt; 1.
DR PRINTS; PR01349; WNTPROTEIN.
DR SMART; SM00097; WNT1; 1.
DR PROSITE; PS00246; WNT1; PARTIAL.
KW Wnt signaling pathway; Developmental protein; Glycoprotein.
FT NON_TER 1
FT TER 134
FT NON_TER 134
FT TER 134
SQ SEQUENCE 134 AA; 14940 MW; 2E7D1E0DCFF01F8B CRC64;

Query Match 0.9%; Score 7; DB 1; Length 134;
Best Local Similarity 100.0%; Pred. No. 22;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 282 RRRGGG 288
DB 123 RRRGGG 129

RESULT 8

YJ01_AQUAE STANDARD; PRT; 135 AA.
ID YJ01_AQUAE AC O67739;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Hypothetical protein AQ_1901.
GN AQ_1901.
OS Aquifex aeolicus.
OC Bacteria; Aquificae; Aquificales; Aquificaceae; Aquifex.
OX NCBI_TaxID=63363;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=VF5;
RX MEDLINE=98196666; PubMed=9537320;
RA Deckert G., Warren P.V., Gaasterland T., Young W.G., Lenox A.L.,
RA Graham D.E., Overbeek R., Snead M.A., Aulay M., Huber R.,
RA Feldman R.A., Short J.M., Olson G.J., Swanson R.V.;
RT "The complete genome of the hyperthermophilic bacterium Aquifex aeolicus".
RL Nature 392:353-358 (1998).
CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).
CC -----

DR EMBL; AE000762; AAC07709.1; -.
DR InterPro; IPR002716; PIN.
DR IntraPro; IPR006596; PIN.
DR Pfam; PF01850; PIN; 1.
DR SMART; SM00670; PIN; 1.
KW Hypothetical protein; Complete proteome.
SQ SEQUENCE 135 AA; 15983 MW; 9C9B90271BFF267D CRC64;

QY 634 FLDCLS 640
DB 97 FLDCLS 103

RESULT 9

VSNI_NEIMC STANDARD; PRT; 148 AA.
ID VSNI_NEIMC AC Q9RLM5;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Putative NmeDIP very-short-patch-repair endonuclease (EC 3.1.-.-) (V.NmeDIP).
GN VSR OR NMEIVP.
OS Neisseria meningitidis (serogroup C).
OC Bacteria; Proteobacteria; Betaproteobacteria; Neisseriales;
OC Neisseriaceae; Neisseria.
OX NCBI_TaxID=135720;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=2120 / Serogroup C / Serotype NT;
RX MEDLINE=20138154; PubMed=10671450;
RA Claus H., Friedrich A., Frosch M., Vogel U.;
RT "Differential distribution of novel restriction-modification systems in clonal lineages of Neisseria meningitidis".
RL J. Bacteriol. 182:1296-1303 (2000).
CC -!- FUNCTION: SPECULATED TO NICK NMEDIP SEQUENCES THAT CONTAIN T:G MISPAAIRS RESULTING FROM M5C-DEAMINATION (BY SIMILARITY).
CC -!- SIMILARITY: BELONGS TO THE VSR FAMILY.
CC -----

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).

DR EMBL; AJ238948; CAB59896.1; -;
 DR HSSP; P09184; 1VSR.
 DR REBASE; 4237; V.NmedIP.
 DR InterPro; IPR004603; Vsr.
 DR Pfam; PF03852; Vsr; 1.
 DR ProDom; PD016088; Vsr; 1.
 KW DNA repair; Hydrolase; Nuclease; Endonuclease; Restriction system.
 SQ SEQUENCE 148 AA; 17419 MW; D6AA99657B48499C CRC64;

Query Match 0.9%; Score 7; DB 1; Length 148;
 Best Local Similarity 100.0%; Pred. No. 24;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 325 RLTPQOR 331
 DB 3 RLTPQOR 9

RESULT 10
 VB15 VACCC STANDARD; PRT; 149 AA.
 ID VB15 VACCC
 AC P21089; 1991 (Rel. 17, Created)
 DT 01-FEB-1991 (Rel. 17, Last sequence update)
 DT 16-OCT-2001 (Rel. 40, Last annotation update)
 DE Protein B15.
 GN B15R.
 OS Vaccinia virus (strain Copenhagen).
 OC Viruses; dsDNA viruses, no RNA stage; Poxviridae; Chordopoxvirinae;
 CC Orthopoxvirus.
 CC NCBI_TaxID=10249;
 RN [1] |||||
 RP SEQUENCE FROM N.A.
 RX MEDLINE=91021027; PubMed=2219722;
 RA Goebel S.J., Johnson G.P., Perkus M.E., Davis S.W., Winslow J.P.,
 RA Paolletti E.;
 RA "The complete DNA sequence of vaccinia virus.";
 RL Virology 179:247-266(1990).
 CC [2]

CC COMPLETE GENOME.
 RA Goebel S.J., Johnson G.P., Perkus M.E., Davis S.W., Winslow J.P.,
 RA Paolletti E.;
 RA "Appendix to 'The complete DNA sequence of vaccinia virus'.";
 RL Virology 179:517-563(1990).
 CC -1- SIMILARITY: TO CAPRIPPOXVIRUS (STRAIN INS-1) AND SHOPE FIBROMA
 CC VIRUS PROTEINS T3A.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).

DR EMBL; M35027; AAA48212.1; -;
 DR PIR; E42527; E42527.
 DR InterPro; IPR003867; Pox_B15.
 DR Pfam; PF02717; Pox_B15; 1.
 DR Early protein.
 KW SEQUENCE 149 AA; 17384 MW; B13DE8D60F671973 CRC64;

Query Match 0.9%; Score 7; DB 1; Length 149;
 Best Local Similarity 100.0%; Pred. No. 24;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 126 YNIFRR 132
 DB 141 YNIFRR 147

RESULT 11
 VB15 VACCC STANDARD; PRT; 149 AA.
 ID VB15 VACCC
 AC P24772;
 DT 01-MAR-1992 (Rel. 21, Created)
 DT 01-MAR-1992 (Rel. 21, Last sequence update)
 DT 01-MAY-1992 (Rel. 22, Last annotation update)
 DE Protein B15.
 GN B15R OR B14R.
 OS Vaccinia virus (strain WR).
 OC Viruses; dsDNA viruses, no RNA stage; Poxviridae; Chordopoxvirinae;
 CC Orthopoxvirus.
 CC NCBI_TaxID=10254;
 RN [1] |||||
 RP SEQUENCE FROM N.A.
 RX MEDLINE=91259063; PubMed=2045793;
 RA Smith G.L., Chan Y.S., Howard S.T.;
 RA "Nucleotide sequence of 42 kbp of vaccinia virus strain WR from near
 RT the right inverted terminal repeat.";
 RL J. Gen. Virol. 72:1349-1376(1991).
 CC [2]

CC SEQUENCE FROM N.A.
 RX MEDLINE=91111982; PubMed=1846491;
 RA Howard S.T., Chan Y.S., Smith G.L.;
 RA "Vaccinia virus homologues of the Shope fibroma virus inverted
 RT terminal repeat proteins and a discontinuous ORF related to the tumor
 RT necrosis factor receptor family.";
 RL Virology 180:633-647(1991).
 CC -1- SIMILARITY: TO CAPRIPPOXVIRUS (STRAIN INS-1) AND SHOPE FIBROMA
 CC VIRUS PROTEINS T3A.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).

DR EMBL; D11079; BAA01844.1; -;
 DR EMBL; M58053; AAA47964.1; -;
 DR PIR; JQ1808; JQ1808.
 DR InterPro; IPR003867; Pox_B15.
 DR Pfam; PF02717; Pox_B15; 1.
 KW Early protein.
 SQ SEQUENCE 149 AA; 17382 MW; A6118A960A0B6973 CRC64;

Query Match 0.9%; Score 7; DB 1; Length 149;
 Best Local Similarity 100.0%; Pred. No. 24;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 126 YNIFRR 132
 DB 141 YNIFRR 147

RESULT 12
 VB15 VARV STANDARD; PRT; 149 AA.
 ID VB15 VARV
 AC P33877;
 DT 01-FEB-1994 (Rel. 28, Created)
 DT 01-FEB-1994 (Rel. 28, Last sequence update)
 DT 01-FEB-1996 (Rel. 33, Last annotation update)
 DE Protein B15.
 GN B15R OR B14R OR B13R.
 OS Variola virus.
 OC Viruses; dsDNA viruses, no RNA stage; Poxviridae; Chordopoxvirinae;

OC Orthopoxvirus.
 OX NCBI_TaxID=10255;
 RN [1]
 RN SEQUENCE FROM N.A.
 RC STRAIN=India-1967 / Isolate Ind3;
 RX MEDLINE=92209372; PubMed=1666548;
 RA Shchelkunov S.N., Marennikova S.S., Totmenin A.V., Blinov V.M.,
 RA Chizhikov V.E., Gutorov V.V., Saifonov P.F., Pozdnyakov S.G.,
 RA Shelukhina E.M., Gashnikov P.V., Anjaparidze O.G., Sandakchchiev L.S.;
 RT "Creation of a clone library of fragments from the natural variola
 virus and study of the structural and functional organization of
 RT viral genes from a circle of hosts";
 RL Dokl. Akad. Nauk SSSR 321:402-406(1991).
 RN [2]
 RN COMPLETE GENOME.
 RP STRAIN=India-1967 / Isolate Ind3;
 RC MEDLINE=93202281; PubMed=8384129;
 RX Shchelkunov S.N., Blinov V.M., Sandakchchiev L.S.;
 RA "Genes of variola and vaccinia viruses necessary to overcome the host
 RT protective mechanisms";
 RL FEBS Lett. 319:80-83(1993).
 RN [3]
 RN SEQUENCE FROM N.A.
 RP STRAIN=Bangladesh-1975;
 RX MEDLINE=94088747; PubMed=8264798;
 RA Massung R.F., Esposito J.J., Liu L., Qi J., Utterback T.R.,
 RA Knight J.C., Aubin L., Yuran T.E., Parsons J.M., Loparev V.N.,
 RA Selivanov N.A., Cavallaro K.F., Kerlavage A.R., Mahy B.W.J.,
 RA Venter C.J.;
 RT "Potential virulence determinants in terminal regions of variola
 RT smallpox virus genome";
 RL Nature 366:748-751(1993).
 RN [4]
 RN SEQUENCE FROM N.A.
 RP STRAIN=Somalia-1977;
 RC STRAIN=Somalia-1977;
 RA Massung R.F., Loparev V.N., Knight J.C., Chizhikov V.E., Parsons J.M.,
 RA Totmenin A.V., Shchelkunov S.N., Esposito J.J.;
 RL Submitted (DEC-1994) to the EMBL/GenBank/DBJ databases.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL; X69198; CAA49123.1; -;
 DR EMBL; L22579; AAA60922.1; -;
 DR EMBL; X67117; CAA47524.1; -;
 DR EMBL; U18341; AAA69454.1; -;
 DR PIR; E36856; E36856.
 DR PIR; T28612; T28612.
 DR InterPro; IPR003867; Pox.B15.
 DR Pfam; PF02717; Pox.B15.1;
 DR SEQUENCE 149 AA; 17369 MW; A84910243C7ED3D4 CRC64;
 Query Match 0.9%; Score 7; DB 1; Length 149;
 Best Local Similarity 100.0%; Pred. No. 24;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 OY 126 YNIIFFR 132
 DB 141 YNIIFFR 147
 RESULT 13
 YFAT_ECOLI
 ID YFAT_ECOLI STANDARD; PRT; 207 AA.
 AC P76456;
 DT 16-OCT-2001 (Rel. 40, Created)
 DT 16-OCT-2001 (Rel. 40, Last sequence update)
 DT 16-OCT-2001 (Rel. 40, Last annotation update)

DE Hypothetical protein yfat precursor.
 GN YFAT OR B2229.
 OS Escherichia coli.
 OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
 OC Enterobacteriaceae; Escherichia.
 OX NCBI_TaxID=562;
 RN [1]
 RN SEQUENCE FROM N.A.
 RP STRAIN=K12 / MG1655;
 RX MEDLINE=97426617; PubMed=9278503;
 RA Blattner F.R., Plunkett G. III, Bloch C.A., Perna N.T., Burland V.,
 RA Riley M., Collado-Vides J., Glasner J.D., Rode C.K., Mayhew G.F.,
 RA Gregor J., Davis N.W., Kirkpatrick H.A., Goeden M.A., Rose D.J.,
 RA Mau B., Shao Y.;
 RT "The complete genome sequence of Escherichia coli K-12";
 RL Science 277:1453-1474(1997).
 CC -!- SIMILARITY: STRONG, TO P.ABRUGINOSA PA4490 AND T.MARITIMA TM0986.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL; AB000312; AAC75289.1; ALT_INIT.
 DR Ecogene; EGI4082; Yfat.
 KW Hypothetical protein; Signal; Complete proteome.
 FT SIGNAL 1 19 POTENTIAL,
 FT CHAIN 20 207 HYPOTHETICAL PROTEIN YFAT.
 SQ SEQUENCE 207 AA; 23767 MW; 6F3386370D78CFCD CRC64;
 Query Match 0.9%; Score 7; DB 1; Length 207;
 Best Local Similarity 100.0%; Pred. No. 32;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 OY 326 LTPEQRQ 332
 DB 91 LTPEQRQ 97
 RESULT 14
 ABFA_PRUPE
 ID ABFA_PRUPE STANDARD; PRT; 209 AA.
 AC Q9ZRA4;
 DT 16-OCT-2001 (Rel. 40, Created)
 DT 16-OCT-2001 (Rel. 40, Last sequence update)
 DT 16-OCT-2001 (Rel. 40, Last annotation update)
 DE Auxin-binding protein ABP19a precursor.
 GS ABP19A.
 OS Prunus persica (Peach).
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
 OC eurosids I; Rosales; Rosaceae; Amygdaloideae; Prunus.
 OX NCBI_TaxID=3760;
 RN [1]
 RN SEQUENCE FROM N.A.
 RP STRAIN=cv. Akatsuki; TISSUE=Shoot apex;
 RX MEDLINE=98329416; PubMed=9664714;
 RA Ohmura A., Tanaka Y., Kadowaki K., Hayashi T.;
 RT "Cloning of genes encoding auxin-binding proteins (ABP19/20) from
 RT peach: significant peptide sequence similarity with germin-like
 RT proteins";
 RL Plant Cell Physiol. 39:492-499(1998).
 CC -!- FUNCTION: PROBABLE RECEPTOR FOR THE PLANT GROWTH-PROMOTING HORMONE
 CC AUXIN.
 CC -!- SUBUNIT: INTERACTS WITH ABP20.
 CC -!- SUBCELLULAR LOCATION: Cell wall.
 CC -!- SIMILARITY: BELONGS TO THE GERMIN FAMILY.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -

CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).

DR EMBL; U79114; AD00295.1; -.
DR HSP; P45850; IFI2.
DR InterPro; IPR006045; Cupin.
DR InterPro; IPR007113; Cupin.sup.
DR InterPro; IPR001929; Germin.
DR Pfam; PF00190; Cupin; 1.
DR PRINTS; PR00325; GERMIN.
DR PROSITE; PS00725; GERMIN; 1.
DR Receptor; Apoplast; Cell wall; Signal; Glycoprotein; Manganese;
KW Metal-binding; Multigene family.
FT SIGNAL 1 18
FT CHAIN 19 209
FT METAL 101 101
FT METAL 103 103
FT METAL 108 108
FT METAL 147 147
FT METAL 147 147
FT DISULFID 24 39
FT CARBOHYD 60 60
SQ SEQUENCE 209 AA; 21859 MW; 886FD412BD8DF771 CRC64;
POTENTIAL.
AUXIN-BINDING PROTEIN ABP19a.
MANGANESE (BY SIMILARITY).
MANGANESE (BY SIMILARITY).
MANGANESE (BY SIMILARITY).
MANGANESE (BY SIMILARITY).
BY SIMILARITY.
N-LINKED (GLCNAC. . .) (POTENTIAL).
SEQUENCE 209 AA; 21859 MW; 886FD412BD8DF771 CRC64;

Query Match 0.9%; Score 7; DB 1; Length 209;
Best Local Similarity 100.0%; Pred. No. 33;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 558 IAGFVAS 564
DB 118 IAGFVAS 124

RESULT 15

ID ABPB_PRUPE STANDARD; PRF; 209 AA.
AC 004012;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Auxin-binding protein ABP19b precursor.
GN ABP19B.
OS Prunus persica (Peach).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
OC eurosids I; Rosales; Rosaceae; Amygdaloideae; Prunus.
OX NCBI_TaxID=3760;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=cv. Akatsuki; TISSUE=Shoot apex;
RX MEDLINE=98329416; PubMed=9664714;
RA Ohmura A., Tanaka Y., Kadowaki K., Hayashi T.;
RT "Cloning of genes encoding auxin-binding proteins (ABP19/20) from
RT peach: significant peptide sequence similarity with germin-like
RT proteins.";
RL Plant Cell Physiol. 39:492-499(1998).
CC -!- FUNCTION: PROBABLE RECEPTOR FOR THE PLANT GROWTH-PROMOTING HORMONE
CC AUXIN.
CC -!- SUBUNIT: INTERACTS WITH ABP20.
CC -!- SUBCELLULAR LOCATION: Cell wall.
CC -!- SIMILARITY: BELONGS TO THE GERMIN FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).

DR EMBL; U01163; AAB51241.1; -.

DR HSP; P45850; IFI2.
DR InterPro; IPR006045; Cupin.
DR InterPro; IPR007113; Cupin.sup.
DR InterPro; IPR001929; Germin.
DR Pfam; PF00190; Cupin; 1.
DR PRINTS; PR00325; GERMIN.
DR PROSITE; PS00725; GERMIN; 1.
DR Receptor; Apoplast; Cell wall; Signal; Glycoprotein; Manganese;
KW Metal-binding; Multigene family.
FT SIGNAL 1 18
FT CHAIN 19 209
FT METAL 101 101
FT METAL 103 103
FT METAL 108 108
FT METAL 147 147
FT METAL 147 147
FT DISULFID 24 39
FT CARBOHYD 60 60
SQ SEQUENCE 209 AA; 21856 MW; 886FD412BD8DF771 CRC64;
POTENTIAL.
AUXIN-BINDING PROTEIN ABP19B.
MANGANESE (BY SIMILARITY).
MANGANESE (BY SIMILARITY).
MANGANESE (BY SIMILARITY).
MANGANESE (BY SIMILARITY).
BY SIMILARITY.
N-LINKED (GLCNAC. . .) (POTENTIAL).
SEQUENCE 209 AA; 21856 MW; 886FD412BD8DF771 CRC64;

Query Match 0.9%; Score 7; DB 1; Length 209;
Best Local Similarity 100.0%; Pred. No. 33;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 558 IAGFVAS 564
DB 118 IAGFVAS 124

Search completed: December 10, 2003, 17:33:55
Job time : 20 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: December 10, 2003, 17:32:31 ; Search time 21 Seconds
(without alignments)
3549.082 Million cell updates/sec

Title: US-10-043-774B-2

Perfect score: 775

Sequence: 1 MIFGVNTRQNLHDHVKESKTG.....VGQSHREPNSLSNRLYYL 775

Scoring table: OLIGO

Gapop 60.0 , Gapext 60.0

Searched: 283308 seqs, 96168682 residues

Word size : 0

Total number of hits satisfying chosen parameters: 283308

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database : PIR 76:*

1: P1:1:*

2: P1:2:*

3: P1:3:*

4: P1:4:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Match	Length	ID	Description
1	15	1.9	722	2 T30995	hypothetical prote
2	15	1.9	824	2 T20351	hypothetical prote
3	9	1.2	958	2 S41013	hypothetical prote
4	9	1.2	1032	2 T23164	hypothetical prote
5	9	1.2	1035	2 T23165	hypothetical prote
6	9	1.2	1040	2 D88568	protein ZK757.3 [i
7	8	1.0	320	2 E64347	hypothetical prote
8	8	1.0	369	2 S13721	Wnt-1 protein prec
9	8	1.0	398	2 T02484	hypothetical prote
10	8	1.0	528	2 T31459	prochlorophyllid
11	8	1.0	1137	2 T19414	hypothetical prote
12	7	0.9	61	2 A31822	hypothetical prote
13	7	0.9	77	2 C71800	hypothetical prote
14	7	0.9	85	2 T44555	hypothetical prote
15	7	0.9	97	2 T31023	conserved hypothe
16	7	0.9	105	2 S27493	nodC protein - Bra
17	7	0.9	112	2 D69543	hypothetical prote
18	7	0.9	130	2 E87024	probable membrane
19	7	0.9	133	2 S6551	YjID protein - Esc
20	7	0.9	133	2 E91289	hypothetical prote
21	7	0.9	133	2 A86131	hypothetical prote
22	7	0.9	134	2 I50729	gene Wnt-1 protein
23	7	0.9	135	2 A10541	probable secreted
24	7	0.9	143	2 T37449	probable 16.7K pro
25	7	0.9	149	2 J01808	B14R 17.3K protein
26	7	0.9	149	2 E42527	B5R protein - vac
27	7	0.9	149	2 F72173	D3R protein - var
28	7	0.9	149	2 E36856	B14R protein - var
29	7	0.9	149	2 T28612	hypothetical prote

30	7	0.9	152	2 H95386	protein imported
31	7	0.9	164	2 T28741	hypothetical prote
32	7	0.9	165	2 T28742	hypothetical prote
33	7	0.9	188	2 B82696	cytochrome B561 XF
34	7	0.9	197	2 T28739	hypothetical prote
35	7	0.9	197	2 S74830	hypothetical prote
36	7	0.9	216	2 C64993	hypothetical prote
37	7	0.9	216	2 H91017	hypothetical prote
38	7	0.9	216	2 B85862	hypothetical prote
39	7	0.9	219	2 F71155	hypothetical prote
40	7	0.9	243	2 C96946	glutamine ABC tran
41	7	0.9	244	1 H64806	YbgL protein - Esc
42	7	0.9	244	2 B90721	probable lactam ut
43	7	0.9	244	2 H85571	probable lactam ut
44	7	0.9	251	2 A83660	hypothetical prote
45	7	0.9	266	2 A69358	carboxylesterase (

ALIGNMENTS

RESULT 1

T30995
hypothetical protein C01G5.2 - Caenorhabditis elegans
C:Species: Caenorhabditis elegans
C:Date: 22-Oct-1999 #sequence_revision 22-Oct-1999 #text_change 29-Oct-1999
C:Accession: T30995
R:Bradshaw, H.; Stellyes, L.
submitted to the EMBL Data Library, August 1999
A:Description: The sequence of C. elegans cosmid C01G5.
A:Reference number: Z20956
A:Accession: T30995
A:Status: preliminary; translated from GB/EMBL/DDBJ
A:Molecule type: DNA
A:Residues: 1-722 <BRA>
A:Cross-references: EMBL:U50068; PIDN:AB37734.1
A:Experimental source: strain Bristol N2
C:Genetics:
A:Map position: IV
A:Introns: 58/3; 121/3; 202/3; 350/1; 471/3; 530/1; 609/2; 671/3
A:Note: C01G5.2

Query Match 1.9%; Score 15; DB 2; Length 722;
Best Local Similarity 100.0%; Pred. No. 1e-06;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY	741	RVPAPCOYAHKLAFL	755
Db	688	RVPAPCOYAHKLAFL	702

RESULT 2

T20351
hypothetical protein D2030.6 - Caenorhabditis elegans
C:Species: Caenorhabditis elegans
C:Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 15-Oct-1999
C:Accession: T20351
R:Wilkinson, J.
submitted to the EMBL Data Library, June 1996
A:Reference number: Z19261
A:Accession: T20351
A:Status: preliminary; translated from GB/EMBL/DDBJ
A:Molecule type: DNA
A:Residues: 1-824 <WTL>
A:Cross-references: EMBL:Z73906; PIDN:CAA98113.1; GSPDB:GN00019; CESP:D2030.6
A:Experimental source: clone D2030
C:Genetics:
A:Gene: CESP:D2030.6
A:Map position: 1
A:Introns: 118/3; 223/3; 304/3; 452/1; 573/3; 632/1; 711/2; 773/3
Query Match 1.9%; Score 15; DB 2; Length 824;
Best Local Similarity 100.0%; Pred. No. 1.1e-06;

Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 741 RVPAPQYAHKLAFL 755
|||||
Db 790 RVPAPQYAHKLAFL 804

RESULT 3

S41013
hypothetical protein ZK757.3 - Caenorhabditis elegans
C:Species: Caenorhabditis elegans
C>Date: 06-Jan-1995 #sequence_revision 06-Jan-1995 #text_change 18-Feb-2000
C:Accession: S41013

R;Thomas, K.
submitted to the EMBL Data Library, December 1993

A:Reference number: S41011
A:Accession: S41013

A>Status: preliminary

A:Molecule type: DNA

A:Residues: 1-958 <THO>

A:Cross-references: EMBL:Z29121

C:Genetics:

A:Introns: 52/2; 350/2; 422/3; 592/1; 679/3; 721/2; 897/2; 929/1

C:Superfamily: rabbit translation initiation factor eIF-2C

Query Match 1.2%; Score 9; DB 2; Length 958;

Best Local Similarity 100.0%; Pred. No. 2.5;

Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 610 RIIVYRDGV 618

Db 729 RIIVYRDGV 737

RESULT 4

T23164
hypothetical protein T22B3.2a - Caenorhabditis elegans
C:Species: Caenorhabditis elegans
C>Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 04-Mar-2000
C:Accession: T23164; T25099

R;Cottage, A.
submitted to the EMBL Data Library, January 1996

A:Reference number: Z19701

A:Accession: T23164

A>Status: preliminary; translated from GB/EMBL/DBDJ

A:Molecule type: DNA

A:Residues: 1-1032 <WIL>

A:Cross-references: EMBL:Z68750; PIDN:CAA92969.1; GSPDB:GN00022; CESP:T22B3.2a

A:Experimental source: clone K01A6

R;Lennard, N.
submitted to the EMBL Data Library, December 1995

A:Reference number: Z19981

A:Accession: T25099

A>Status: preliminary; translated from GB/EMBL/DBDJ

A:Molecule type: DNA

A:Residues: 1-1032 <WIL>

A:Cross-references: EMBL:Z68300; PIDN:CAA92618.1; GSPDB:GN00022; CESP:T22B3.2a

A:Experimental source: clone T22B3

C:Genetics:

A:Gene: CESP:T22B3.2a

A:Map position: 4

A:Introns: 19/1; 71/3; 129/2; 424/2; 496/3; 607/1; 666/1; 753/3; 795/2; 961/2; 1003/1

C:Superfamily: rabbit translation initiation factor eIF-2C

Query Match 1.2%; Score 9; DB 2; Length 1032;

Best Local Similarity 100.0%; Pred. No. 2.7;

Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 610 RIIVYRDGV 618

Db 803 RIIVYRDGV 811

RESULT 5

T23165

hypothetical protein T22B3.2b - Caenorhabditis elegans

C:Species: Caenorhabditis elegans

C>Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 04-Mar-2000

C:Accession: T23165; T25100

R;Cottage, A.
submitted to the EMBL Data Library, January 1996

A:Reference number: Z19701

A:Accession: T23165

A>Status: preliminary; translated from GB/EMBL/DBDJ

A:Molecule type: DNA

A:Residues: 1-1035 <WIL>

A:Cross-references: EMBL:Z68750; PIDN:CAA92970.1; GSPDB:GN00022; CESP:T22B3.2b

A:Experimental source: clone K01A6

R;Lennard, N.
submitted to the EMBL Data Library, December 1995

A:Reference number: Z19981

A:Accession: T25100

A>Status: preliminary; translated from GB/EMBL/DBDJ

A:Molecule type: DNA

A:Residues: 1-1035 <WIL>

A:Cross-references: EMBL:Z68300; PIDN:CAA92619.1; GSPDB:GN00022; CESP:T22B3.2b

A:Experimental source: clone T22B3

C:Genetics:

A:Gene: CESP:T22B3.2b

A:Map position: 4

A:Introns: 19/1; 74/3; 132/2; 427/2; 499/3; 610/1; 669/1; 756/3; 798/2; 964/2; 1006/1

C:Superfamily: rabbit translation initiation factor eIF-2C

Query Match 1.2%; Score 9; DB 2; Length 1035;

Best Local Similarity 100.0%; Pred. No. 2.7;

Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 610 RIIVYRDGV 618

Db 806 RIIVYRDGV 814

RESULT 6

D88568
protein ZK757.3 [imported] - Caenorhabditis elegans

C:Species: Caenorhabditis elegans

C>Date: 10-May-2001 #sequence_revision 10-May-2001 #text_change 24-May-2001

C:Accession: D88568

R;anonymous, The C. elegans Sequencing Consortium.

Science 282, 2012-2018, 1998

A:Title: Genome sequence of the nematode C. elegans: a platform for investigating biol

A:Reference number: A75000; MUID:99069613; PMID:9851916

A>Note: see websites genome.wustl.edu/gsc/C_elegans/ and www.sanger.ac.uk/projects/C_e

A>Note: published errata appeared in Science 283, 35, 1999; Science 283, 2103, 1999; a

A:Accession: D88568

A>Status: preliminary

A:Molecule type: DNA

A:Residues: 1-1040 <STO>

A:Cross-references: GB:chr_III; PIDN:CAA82941.1; PID:93877004; GSPDB:GN00021; CESP:ZK7

C:Genetics:

A:Gene: ZK757.3

A:Map position: 3

C:Superfamily: rabbit translation initiation factor eIF-2C

Query Match 1.2%; Score 9; DB 2; Length 1040;

Best Local Similarity 100.0%; Pred. No. 2.7;

Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 610 RIIVYRDGV 618

Db 811 RIIVYRDGV 819

RESULT 7

B64347

hypothetical protein MJ0381 - Methanococcus jannaschii

C;Species: Methanococcus jannaschii
 C;Date: 13-Sep-1996 #sequence_revision 13-Sep-1996 #text_change 21-Jul-2000
 C;Accession: E64347
 R;Bult, C.J.; White, O.; Olsen, G.J.; Zhou, L.; Fleischmann, R.D.; Sutton, G.G.; Blake, R.; Reich, C.I.; Overbeek, R.; Kirkness, E.F.; Weinstock, K.G.; Merrick, J.M.; Glodex, A.; Raun, J.D.; Sadow, P.W.; Hanna, M.C.; Cotton, M.D.; Roberts, K.M.; Hurst, M.A. Science 273, 1058-1073, 1996
 A;Authors: Kaine, B.P.; Borodovsky, M.; Klenk, H.P.; Fraser, C.M.; Smith, H.O.; Woese, C. A;Title: Complete genome sequence of the methanogenic archaeon, Methanococcus jannaschii A;Reference number: A64300; MUID:96337999; PMID:8688087
 A;Accession: E64347
 A;Status: preliminary; nucleic acid sequence not shown; translation not shown
 A;Molecule type: DNA
 A;Residues: 1-320 <BUL>
 A;Cross-references: GB:U067491; GB:L77117; NID:g2826278; PIDN:AAB98370.1; PID:g1591087; T C;Genetics:
 A;Map position: FOR345977-346939
 C;Superfamily: Pyrococcus horikoshii hypothetical protein PH0920

Query Match 1.0%; Score 8; DB 2; Length 320;
 Best Local Similarity 100.0%; Pred. No. 10;
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 247 KKADGSEV 254
 |||||
 DB 95 KKADGSEV 102

RESULT 8
 S13721
 Wnt-1 protein precursor, secreted - axolotl
 C;Species: Ambystoma mexicanum (axolotl)
 C;Date: 21-Nov-1993 #sequence_revision 10-Nov-1995 #text_change 16-Jul-1999
 C;Accession: S13721; S13722
 R;Busse, U.; Guay, J.; Seguin, C. Nucleic Acids Res. 18, 7439, 1990
 A;Title: Nucleotide sequence of a cDNA encoding Wnt-1 of the Mexican axolotl Ambystoma A;Reference number: S13721; MUID:91081334; PMID:2259633
 A;Accession: S13721
 A;Status: preliminary; nucleic acid sequence not shown; translation not shown
 A;Molecule type: mRNA
 A;Residues: 1-369 <BUS>
 A;Cross-references: EMBL:X55270; NID:g62424; PIDN:CAA38991.1; PID:g62425
 A;Note: the nucleotide sequence was submitted to the EMBL Data Library, October 1990
 R;Busse, U.; Guay, J.; Seguin, C. Nucleic Acids Res. 19, 981, 1991
 A;Title: Nucleotide sequence of a cDNA encoding Wnt-1 of the Mexican axolotl Ambystoma A;Reference number: S13722; MUID:91204483; PMID:2017393
 A;Contents: annotation; reprinted sequence figure
 C;Genetics:
 A;Gene: Wnt-1
 C;Superfamily: int-1 transforming protein

Query Match 1.0%; Score 8; DB 2; Length 369;
 Best Local Similarity 100.0%; Pred. No. 12;
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 282 RRRPGPGT 289
 |||||
 DB 155 RRRPGPGT 162

RESULT 9
 T02484
 hypothetical protein At2g30010 [imported] - Arabidopsis thaliana
 N;Alternate names: hypothetical protein F23F1.7
 C;Species: Arabidopsis thaliana (mouse-ear cress)
 C;Date: 05-Mar-1999 #sequence_revision 05-Mar-1999 #text_change 16-Feb-2001
 C;Accession: T02484; D84703
 R;Rounsley, S.D.; Lin, X.; Ketchum, K.A.; Crosby, M.L.; Brandon, R.C.; Sykes, S.M.; Kaul submitted to the EMBL Data Library, August 1998
 A;Description: Arabidopsis thaliana chromosome II BAC F23F1 genomic sequence.
 A;Reference number: Z14675

A;Accession: T02484
 A;Status: translated from GB/EMBL/DBJ
 A;Molecule type: DNA
 A;Residues: 1-398 <ROU>
 A;Cross-references: EMBL:AC004680; NID:g3420043; PID:g3420050
 A;Experimental source: cultivar Columbia
 R;Lin, X.; Kaul, S.; Rounsley, S.D.; Shea, T.P.; Benito, M.I.; Town, C.D.; Fujii, C.Y. M.; Koo, H.; Moffat, K.S.; Cronin, L.A.; Shen, M.; Vanaken, S.E.; Umayam, L.; Tallon, euss, D.; Nierman, W.C.; White, O.; Eisen, J.A.; Salzberg, S.L.; Fraser, C.M.; Venter, Nature 402, 761-768, 1999
 A;Title: Sequence and analysis of chromosome 2 of the plant Arabidopsis thaliana.
 A;Reference number: A84420; MUID:20083487; PMID:10617197
 A;Accession: D84703
 A;Status: preliminary
 A;Molecule type: DNA
 A;Residues: 1-398 <STO>
 A;Cross-references: GB:AE002093; NID:g3420050; PIDN:AAC31851.1; GSPDB:GN00139 C;Genetics:
 A;Gene: At2G30010; F23F1.7
 A;Map position: 2
 A;Introns: 112/2; 169/3; 230/2; 283/2

Query Match 1.0%; Score 8; DB 2; Length 398;
 Best Local Similarity 100.0%; Pred. No. 13;
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 702 SVSPHYN 709
 |||||
 DB 276 SVSPHYN 283

RESULT 10
 T31459
 protochlorophyllide reductase (EC 1.3.1.33) chain B BchB, light-independent - Helioobac C;Species: Helioobacillus mobilis
 C;Date: 02-Sep-2000 #sequence_revision 02-Sep-2000 #text_change 15-Sep-2000
 C;Accession: T31459
 R;Xiong, J.; Inoue, K.; Bauer, C.E. Proc. Natl. Acad. Sci. U.S.A. 95, 14851-14856, 1998
 A;Title: Tracking molecular evolution of photosynthesis by characterization of a major A;Reference number: Z21036; MUID:99061957; PMID:9843979
 A;Accession: T31459
 A;Status: preliminary; translated from GB/EMBL/DBJ
 A;Molecule type: DNA
 A;Residues: 1-528 <XIO>
 A;Cross-references: EMBL:AF080002; NID:g3820536; PID:g3820557; PIDN:AAC84030.1 C;Genetics:
 A;Gene: bchB
 C;Superfamily: protochlorophyllide reductase chain chlB
 C;Keywords: oxidoreductase

Query Match 1.0%; Score 8; DB 2; Length 528;
 Best Local Similarity 100.0%; Pred. No. 16;
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 501 VARTLGKQ 508
 |||||
 DB 115 VARTLGKQ 122

RESULT 11
 T19414
 hypothetical protein C23H4.6 - Caenorhabditis elegans
 C;Species: Caenorhabditis elegans
 C;Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 18-Feb-2000
 C;Accession: T19414
 R;Wilkinson, J. submitted to the EMBL Data Library, August 1996
 A;Reference number: Z19121
 A;Accession: T19414
 A;Status: preliminary; translated from GB/EMBL/DBJ
 A;Molecule type: DNA
 A;Residues: 1-1137 <WIL>

A;Cross-references: EMBL:Z78416; PIDN:CA801681.1; GSPDB:GN00028; CBSP:C23H4.6

A;Experimental source: clone C23H4

C;Genetics:

A;Gene: CBSP:C23H4.6

A;Map position: X

A;Introns: 34/1; 76/1; 111/2; 482/2; 598/2; 1091/3

Query Match 1.0%; Score 8; DB 2; Length 1137;

Best Local Similarity 100.0%; Pred. No. 32;

Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 107 ITITLTNE 114

DB 115 ITITLTNE 122

RESULT 12

AG1862

hypothetical protein asl0448 [imported] - Nostoc sp. (strain PCC 7120)

C;Species: Nostoc sp. PCC 7120

A;Note: Nostoc sp. strain PCC 7120 is a synonym of Anabaena sp. strain PCC 7120

C;Date: 14-Dec-2001 #sequence_revision 14-Dec-2001 #text_change 09-Dec-2002

C;Accession: AG1862

R;Kaneko, T.; Nakamura, Y.; Wolk, C.P.; Kuritz, T.; Sasamoto, S.; Watanabe, A.; Iriiguchi

Nakazaki, N.; Shimo, S.; Sugimoto, M.; Takazawa, M.; Yanada, M.; Yasuda, M.; Tabata, S.

DNA Res. 8, 205-213, 2001

A;Title: Complete Genomic Sequence of the Filamentous Nitrogen-fixing Cyanobacterium Ana

A;Reference number: AB1807; MUID:21595285; PMID:11759840

A;Accession: AG1862

A;Status: preliminary

A;Molecule type: DNA

A;Residues: 1-61 <KUR>

A;Cross-references: GB:BA000019; PIDN:BA872406.1; PID:gl7129793; GSPDB:GN00179

A;Experimental source: strain PCC 7120

C;Genetics:

A;Gene: asl0448

Query Match

Best Local Similarity 0.9%; Score 7; DB 2; Length 61;

Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 512 MAIATKI 518

DB 36 MAIATKI 42

RESULT 13

C71800

hypothetical protein jhp1495 - Helicobacter pylori (strain J99)

C;Species: Helicobacter pylori

A;Variety: strain J99

C;Date: 12-Feb-1999 #sequence_revision 12-Feb-1999 #text_change 08-Oct-1999

C;Accession: C71800

R;Alm, R.A.; Ling, L.S.L.; Moir, D.T.; King, B.L.; Brown, E.D.; Doig, P.C.; Smith, D.R.

Ives, C.; Gibson, R.; Merberg, D.; Mills, S.D.; Jiang, Q.; Taylor, D.E.; Vovis, G.F.;

Nature 397, 178-180, 1999

A;Title: Genomic sequence comparison of two unrelated isolates of the human gastric path

A;Reference number: A71800; MUID:99120557; PMID:9923682

A;Accession: C71800

A;Status: preliminary

A;Molecule type: DNA

A;Residues: 1-77 <ARN>

A;Cross-references: GB:AE001571; GB:AE001439; NID:g4156120; PIDN:AA07071.1; PID:g415612

A;Experimental source: strain J99

C;Genetics:

A;Gene: jhp1495

Query Match

Best Local Similarity 0.9%; Score 7; DB 2; Length 77;

Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 217 LIGLVVL 223

DB 115 LIGLVVL 122

DB 63 LIGLVVL 69

RESULT 14

T44555

hypothetical protein PA0631 [imported] - Pseudomonas aeruginosa (strain PA01)

C;Species: Pseudomonas aeruginosa

C;Date: 31-Jan-2000 #sequence_revision 31-Jan-2000 #text_change 06-Oct-2000

C;Accession: T44555; E83568

R;Nakayama, K.; Takashima, K.; Ishihara, H.; Shinomiya, T.; Kageyama, M.; Kanaya, S.;

submitted to the EMBL Data Library, August 1999

A;Description: Genetic relationship between bacteriophages and bacteriophages.

A;Reference number: Z22790

A;Accession: T44555

A;Status: preliminary; translated from GB/EMBL/DDBJ

A;Molecule type: DNA

A;Residues: 1-85 <NAK>

A;Cross-references: EMBL:AB030825; PIDN:BA883170.1

A;Experimental source: strain PA01

R;Stover, C.K.; Pham, X.Q.; Erwin, A.L.; Mizoguchi, S.D.; Warrenner, P.; Hickey, M.J.;

Adman, S.; Yuan, Y.; Brody, L.F.; Coulter, S.N.; Folger, K.R.; Kas, A.; Larbig, K.; Li

; Lory, S.; Olson, M.V.

Nature 406, 959-964, 2000

A;Title: Complete genome sequence of Pseudomonas aeruginosa PA01, an opportunistic pat

A;Reference number: AB2950; MUID:20437337; PMID:10984043

A;Accession: E83568

A;Status: preliminary

A;Molecule type: DNA

A;Residues: 1-85 <STO>

A;Cross-references: GB:AE004498; GB:AE004091; NID:g9946491; PIDN:AAG04020.1; GSPDB:GN

A;Experimental source: strain PA01

C;Genetics:

A;Gene: PA0631

Query Match

Best Local Similarity 0.9%; Score 7; DB 2; Length 85;

Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 593 LQAALRA 599

DB 77 LQAALRA 83

RESULT 15

T31023

conserved hypothetical protein 108 - Sulfolobus sp. plasmid pNOB8

C;Species: Sulfolobus sp.

C;Date: 02-Sep-2000 #sequence_revision 02-Sep-2000 #text_change 02-Sep-2000

C;Accession: T31023

R;She, Q.; Phan, H.; Garrett, R.A.; Albers, S.V.; Stedman, K.M.; Zillig, W.

Extremophiles 2, 417-425, 1998

A;Title: Genetic profile of pNOB8 from Sulfolobus: the first conjugative plasmid from

A;Reference number: Z20959; MUID:99044580; PMID:9827331

A;Accession: T31023

A;Status: preliminary; translated from GB/EMBL/DDBJ

A;Molecule type: DNA

A;Residues: 1-97 <SHB>

A;Cross-references: EMBL:AU010405; NID:el351926; PID:el351945; PIDN:CAA09129.1

A;Experimental source: strain NOB8H2

C;Genetics:

A;Genome: plasmid pNOB8

Query Match

Best Local Similarity 0.9%; Score 7; DB 2; Length 97;

Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 679 IDVEVTR 685

DB 60 IDVEVTR 66

Search completed: December 10, 2003, 17:35:29

Job time : 24 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: December 10, 2003, 17:32:26 ; Search time 47 Seconds
(without alignments)
2617.300 Million cell updates/sec

Title: US-10-043-774B-2
Perfect score: 775
Sequence: 1 MIFGVTNRQLDHVKESKTG.....VQGSIHREPULSLNRLYYL 775

Scoring table:

Gapop 60.0 , Gapext 60.0

Searched: 1107863 seqs, 158726573 residues

Word size : 0

Total number of hits satisfying chosen parameters: 1107863

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database : A_Geneseq_19Jun03.*
1: /SIDSL1/gcgdata/geneseq/geneseq-emb1/AA1980.DAT.*
2: /SIDSL1/gcgdata/geneseq/geneseq-emb1/AA1981.DAT.*
3: /SIDSL1/gcgdata/geneseq/geneseq-emb1/AA1982.DAT.*
4: /SIDSL1/gcgdata/geneseq/geneseq-emb1/AA1983.DAT.*
5: /SIDSL1/gcgdata/geneseq/geneseq-emb1/AA1984.DAT.*
6: /SIDSL1/gcgdata/geneseq/geneseq-emb1/AA1985.DAT.*
7: /SIDSL1/gcgdata/geneseq/geneseq-emb1/AA1986.DAT.*
8: /SIDSL1/gcgdata/geneseq/geneseq-emb1/AA1987.DAT.*
9: /SIDSL1/gcgdata/geneseq/geneseq-emb1/AA1988.DAT.*
10: /SIDSL1/gcgdata/geneseq/geneseq-emb1/AA1989.DAT.*
11: /SIDSL1/gcgdata/geneseq/geneseq-emb1/AA1990.DAT.*
12: /SIDSL1/gcgdata/geneseq/geneseq-emb1/AA1991.DAT.*
13: /SIDSL1/gcgdata/geneseq/geneseq-emb1/AA1992.DAT.*
14: /SIDSL1/gcgdata/geneseq/geneseq-emb1/AA1993.DAT.*
15: /SIDSL1/gcgdata/geneseq/geneseq-emb1/AA1994.DAT.*
16: /SIDSL1/gcgdata/geneseq/geneseq-emb1/AA1995.DAT.*
17: /SIDSL1/gcgdata/geneseq/geneseq-emb1/AA1996.DAT.*
18: /SIDSL1/gcgdata/geneseq/geneseq-emb1/AA1997.DAT.*
19: /SIDSL1/gcgdata/geneseq/geneseq-emb1/AA1998.DAT.*
20: /SIDSL1/gcgdata/geneseq/geneseq-emb1/AA1999.DAT.*
21: /SIDSL1/gcgdata/geneseq/geneseq-emb1/AA2000.DAT.*
22: /SIDSL1/gcgdata/geneseq/geneseq-emb1/AA2001.DAT.*
23: /SIDSL1/gcgdata/geneseq/geneseq-emb1/AA2002.DAT.*
24: /SIDSL1/gcgdata/geneseq/geneseq-emb1/AA2003.DAT.*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Query Match	Score	Length	ID	Description
1	522	67.4	523	23	ABP67995 Human colon cancer
2	421	54.3	861	21	AA190235 Human piwi protein
3	74	9.5	862	21	AA190234 Mouse piwi protein
4	67	8.6	130	21	AA190235 Breast and ovarian
5	11	1.4	498	23	AA190235 Human mRNA sequenc
6	9	1.2	866	22	AA190235 Drosophila melanog
7	8	1.0	71	23	AA190235 Human peptide enco
8	8	1.0	140	22	AA190235 Novel human diagn
9	8	1.0	254	21	AA190235 Arabidopsis thalia

10 8 1.0 271 21 AAG20039 Arabidopsis thalia
11 8 1.0 277 21 AAG20038 Arabidopsis thalia
12 8 1.0 277 21 AAG48143 Arabidopsis thalia
13 8 1.0 371 22 AAB92722 Human protein sequ
14 8 1.0 398 21 AAG48142 Arabidopsis thalia
15 8 1.0 433 21 AAG48141 Arabidopsis thalia
16 8 1.0 530 22 AAB94209 Human protein sequ
17 8 1.0 580 22 AAU07866 Polypeptide sequen
18 7 0.9 54 22 AAM85210 Human immune/haema
19 7 0.9 55 22 ABB37543 Peptide #5049 enco
20 7 0.9 55 22 ABE22837 Protein #4836 enco
21 7 0.9 55 23 ABG40342 Human peptide enco
22 7 0.9 60 22 AAU47180 Propionibacterium
23 7 0.9 60 22 AAM30064 Human immune/haema
24 7 0.9 68 22 ABB22884 Escherichia coli p
25 7 0.9 74 20 AAU02975 Fragment of human
26 7 0.9 75 21 AAY56453 Neisseria gonorrhe
27 7 0.9 76 22 AAM99697 Human excretory re
28 7 0.9 76 22 AAM42512 Human kidney relat
29 7 0.9 77 22 AAU65120 Propionibacterium
30 7 0.9 78 23 ABP06411 Human ORFX protein
31 7 0.9 84 22 ABG02717 Novel human diagn
32 7 0.9 91 24 ABB58203 Human nervous syst
33 7 0.9 99 20 AAY02974 Rice stress respon
34 7 0.9 101 18 AAW20867 Human ORFX protein
35 7 0.9 109 21 AAG04298 Arabidopsis thalia
36 7 0.9 111 21 AAG04297 Arabidopsis thalia
37 7 0.9 120 21 AAB40477 Human ORFX ORF241
38 7 0.9 130 23 ABU05453 M. tuberculosis an
39 7 0.9 146 23 ABB89863 Human polypeptide
40 7 0.9 164 22 AAG66996 Human phosphorylas
41 7 0.9 164 22 AAB80615 Environmental stre
42 7 0.9 165 21 AAB40995 Human ORFX ORF759
43 7 0.9 169 22 AAO13144 Human polypeptide
44 7 0.9 176 23 ABP33797 Staphylococcus epi
45 7 0.9 176 23 ABP33797

ALIGNMENTS

RESULT 1
ABP67995
ID ABP67995 standard; Protein; 523 AA.
XX
AC ABP67995;
XX
DT 13-DEC-2002 (first entry)
XX
DE Human colon cancer related polypeptide SEQ ID NO 2603.
XX
KW Human; colon; cancer; cytostatic; tumour; gene therapy; vaccine.
XX
OS Homo sapiens.
XX
PN WO200258534-A2.
XX
PD 01-AUG-2002.
XX
PF 19-NOV-2001; 2001WO-US43704.
XX
PR 20-NOV-2000; 2000US-252222P.
XX
PR 06-FEB-2001; 2001US-267011P.
XX
PR 28-MAR-2001; 2001US-279670P.
XX
PR 10-JUL-2001; 2001US-304037P.
XX
XX (CORI-) CORIXA CORP.
XX
XX Stolk JA, Xu J; Chenault RA, Meagher MJ, Secrist H, King GE;
XX
XX WPI; 2002-608400/65.
XX
XX N-ESDB; ABV89287.

PT New isolated tumor colon polynucleotide and polypeptide, useful for the
 PT diagnosis, prevention and/or treatment of cancer, in particular colon
 PT cancer

XX Claim 2: SEQ ID NO 2603; 266pp + Sequence Listing; English.

XX The invention relates to a human colon tumour expressed polynucleotide
 CC (i) encoding a polypeptide (ii), ABP67991-ABP67996) comprising: (i) any of
 CC 2600 fully defined nucleotide sequences (ABV8669-ABV89289); (ii)
 CC complements of (i); (iii) at least 20 contiguous residues of (i); (iv)
 CC sequences that hybridize to (i), under moderately stringent conditions;
 CC (v) sequences having at least 75% or 90% identity to (i); or (vi)
 CC degenerate variants of (i). The compositions and methods of the present
 CC invention are useful for the diagnosis, prevention and/or treatment of
 CC cancer, particularly colon cancer. (i) can be used in gene therapy and
 CC (ii) and (iii) are useful in pharmaceutical compositions such as vaccines.
 CC Note: The sequence data for this patent did not form part of the printed
 CC specification, but was obtained in electronic format directly from WIPO
 CC at ftp.wipo.int/pub/published_pct_sequences.

XX Sequence 523 AA;

Query Match 67.4%; Score 522; DB 23; Length 523;
 Best Local Similarity 100.0%; Pred. No. 0;
 Matches 522; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 254 VSLEYRKYQNOEITDLKQPVLSQPKRRGGTLPGPAMLIPELCYLTGLTDKWRND 313
 DB 2 VSLEYRKYQNOEITDLKQPVLSQPKRRGGTLPGPAMLIPELCYLTGLTDKWRND 61

QY 314 FNYMKDLAVHTRLTPEQORQEVGRGLDYLTHKNDNVQRELDRWGLSPDSNLLSFGRILOT 373
 DB 62 FNYMKDLAVHTRLTPEQORQEVGRGLDYLTHKNDNVQRELDRWGLSPDSNLLSFGRILOT 121

QY 374 EKHQGGKTFDYNPQFADWSKETRGAPLISVKPLDNWLLIYTRNVEAANSLIQNLFKVT 433
 DB 122 EKHQGGKTFDYNPQFADWSKETRGAPLISVKPLDNWLLIYTRNVEAANSLIQNLFKVT 181

QY 434 PANGMQMRKAIMIEVDRTAYLRVLQOKVTADTQIVVCLLSNRKDKYDAIKKYLCTDC 493
 DB 182 PANGMQMRKAIMIEVDRTAYLRVLQOKVTADTQIVVCLLSNRKDKYDAIKKYLCTDC 241

QY 494 PTPSQCVVARTLGKQQTVMIAITKIALQNMCKMGELWRVDIPLKLVMIIVGIDYHDMTA 553
 DB 242 PTPSQCVVARTLGKQQTVMIAITKIALQNMCKMGELWRVDIPLKLVMIIVGIDYHDMTA 301

QY 554 GRSIAGFVASINEGMRTRWFSRCIFQDRGOELVDGLKVCLOAALRAWNSCNEYMPSRRIIV 613
 DB 302 GRSIAGFVASINEGMRTRWFSRCIFQDRGOELVDGLKVCLOAALRAWNSCNEYMPSRRIIV 361

QY 614 YRDGVGGQKLTLVNVEVPQFDCCKSIGRGYNPRLTVIVVKRVNTRFPAQSGRLQNP 673
 DB 362 YRDGVGGQKLTLVNVEVPQFDCCKSIGRGYNPRLTVIVVKRVNTRFPAQSGRLQNP 421

QY 674 LPGVIVDEVRPEWYDFFIVSQAVRSGSVSPTHYVINDNSGLKPDHIORLTYKLCHYI 733
 DB 422 LPGVIVDEVRPEWYDFFIVSQAVRSGSVSPTHYVINDNSGLKPDHIORLTYKLCHYI 481

QY 734 YNPFQVIRVPAPCOYAHKLAFLVQGSIHREPNLSNRLYYL 775
 DB 482 YNPFQVIRVPAPCOYAHKLAFLVQGSIHREPNLSNRLYYL 523

RESULT 2

AAV90235

ID AAV90235 standard; Protein; 861 AA.

XX

AC AAV90235;

XX

DT 29-AUG-2000 (first entry)

XX

DE Human piwi protein, designated hiwi.

XX

KW Piwi family protein; piwi; miwi; hiwi; gene therapy; tissue dystrophy;
 KW anaemia; immunodeficiency; male infertility; human.

OS Homo sapiens.

XX Location/Qualifiers
 PH Key
 FT Misc-difference 76
 FT /label= Leu, Ile
 FT /note= "encoded by NTA"
 FT Misc-difference 303
 FT /label= Leu, Ile
 FT /note= "encoded by NTA"
 FT Misc-difference 735
 FT /label= Leu, Ile
 FT /note= "encoded by NTA"

XX WO200032039-A1.

XX 08-JUN-2000.

XX 03-DEC-1999; 99WO-US28764.

XX 04-DEC-1998; 98US-0110901.

XX (UYDU-) UNIV DUKE.

XX Lin H;

XX WPI; 2000-412085/35.

XX N-PSDB; AAA07588.

XX Piwi family nucleic acids, polypeptides, and antibodies, useful in gene
 PT therapy of diseases such as cancer and in various research and
 PT diagnostic applications

XX Claim 4; Page 189-194; 201pp; English.

XX This sequence represents the human piwi family protein, designated
 CC hiwi. The piwi family nucleic acids and polypeptides are used in gene
 CC therapy of diseases such as cancer and also in various research and
 CC diagnostic applications. The sequences can also be used to treat
 CC tissue dystrophy, anaemia, immunodeficiency, and male infertility.

XX Sequence 861 AA;

Query Match 54.3%; Score 421; DB 21; Length 861;
 Best Local Similarity 99.8%; Pred. No. 0;
 Matches 521; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 254 VSLEYRKYQNOEITDLKQPVLSQPKRRGGTLPGPAMLIPELCYLTGLTDKWRND 313

DB 340 VSLEYRKYQNOEITDLKQPVLSQPKRRGGTLPGPAMLIPELCYLTGLTDKWRND 399

QY 314 FNYMKDLAVHTRLTPEQORQEVGRGLDYLTHKNDNVQRELDRWGLSPDSNLLSFGRILOT 373

DB 400 FNYMKDLAVHTRLTPEQORQEVGRGLDYLTHKNDNVQRELDRWGLSPDSNLLSFGRILOT 459

QY 374 EKHQGGKTFDYNPQFADWSKETRGAPLISVKPLDNWLLIYTRNVEAANSLIQNLFKVT 433

DB 460 EKHQGGKTFDYNPQFADWSKETRGAPLISVKPLDNWLLIYTRNVEAANSLIQNLFKVT 519

QY 434 PANGMQMRKAIMIEVDRTAYLRVLQOKVTADTQIVVCLLSNRKDKYDAIKKYLCTDC 493

DB 520 PANGMQMRKAIMIEVDRTAYLRVLQOKVTADTQIVVCLLSNRKDKYDAIKKYLCTDC 579

QY 494 PTPSQCVVARTLGKQQTVMIAITKIALQNMCKMGELWRVDIPLKLVMIIVGIDYHDMTA 553

DB 580 PTPSQCVVARTLGKQQTVMIAITKIALQNMCKMGELWRVDIPLKLVMIIVGIDYHDMTA 639

QY 554 GRSIAGFVASINEGMRTRWFSRCIFQDRGOELVDGLKVCLOAALRAWNSCNEYMPSRRIIV 613

DB 640 GRSIAGFVASINEGMRTRWFSRCIFQDRGOELVDGLKVCLOAALRAWNSCNEYMPSRRIIV 699

SQ Sequence 130 AA;

Query Match 8.6%; Score 67; DB 21; Length 130;
 Best Local Similarity 100.0%; Pred. No. 8.4e-61; DB 21; Length 130;
 Matches 67; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 549 HDMTAGRSIAGFVASINEGMTWFSRCIFQDGGELVDGLKVCLOALRAWNSCNEYMP 608
 |||||
 DB 62 HDMTAGRSIAGFVASINEGMTWFSRCIFQDGGELVDGLKVCLOALRAWNSCNEYMP 121

QY 609 SRILVVR 615
 |||||
 DB 122 SRILVVR 128

RESULT 5

ABP43865
 ID ABP43865 standard; Protein; 498 AA.
 AC ABP43865;
 DT 26-FEB-2003 (first entry)
 DE Human mRNA sequence.
 KW Neuroprotective; immunomodulator; cancer;
 KW cytosolic; anti-inflammatory; gene therapy; nutritional supplement;
 KW wound; burn; ulcer; Alzheimer's disease; Huntington's disease;
 KW amyotrophic lateral sclerosis; autoimmune disorder; inflammation;
 KW vulvectomy.

OS Homo sapiens.

XX WO200231111-A2.

PN 18-APR-2002.

PD 11-OCT-2001; 2001WO-US27760.

PF 12-OCT-2000; 2000US-0687527.

PR (HYSE-) HYSEQ INC.

XX Tang YT, Liu C, Zhou P, Asundi V, Zhang J, Zhao QA, Ren F;

XX Xue AJ, Yang Y, Wehrman T, Drmanac RT;

XX WPI; 2002-426278/45.

XX N-PSDB; ABQ61109.

XX New polypeptides and their encoded proteins, useful as nutritional

XX sources or supplements, or in gene therapy, particularly for treating

XX wounds, Alzheimer's disease, amyotrophic lateral sclerosis, cancer or

XX inflammation

XX Claim 20; SEQ ID # 768; 357pp + sequence listing; English.

XX The invention relates to 446 newly isolated polynucleotide sequences.

XX The activity of polynucleotides of the invention may be described as,

XX vulnary, neuroprotective, immunomodulator, cytosolic and

XX anti-inflammatory. Compositions comprising nucleic acids of the invention

XX are useful for treating a mammalian subject, or as nutritional sources or

XX supplements. These are useful in gene therapy, particularly for treating

XX wounds, burns or ulcers, Alzheimer's disease, Huntington's disease,

XX amyotrophic lateral sclerosis, autoimmune disorders, cancer or

XX inflammation. The nucleic acids and polypeptides are also useful in

XX diagnostic and research methods. The sequences given in records

XX ABP43844-ABP43889 represent polypeptides encoded by polynucleotides of

XX the invention.

XX NOTE: The sequence data for this patent did not form part of the printed

XX specification, but was obtained in electronic format directly from WIPO

XX at ftp.wipo.int/pub/published_pct_sequences.

XX Sequence 498 AA;

Query Match

1.4%; Score 11; DB 23; Length 498;
 Best Local Similarity 100.0%; Pred. No. 0.042;
 Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 742 VPAPCOYAHKL 752

DB 465 VPAPCOYAHKL 475

RESULT 6

ABB62102
 ID ABB62102 standard; Protein; 866 AA.
 AC ABB62102;
 DT 26-MAR-2002 (first entry)
 DE Drosophila melanogaster polypeptide SEQ ID NO 13098.
 KW Drosophila; developmental biology; cell signalling; insecticide;
 KW pharmaceutical.
 KW Drosophila melanogaster.
 OS Drosophila melanogaster.
 WO200171042-A2.
 27-SEP-2001.
 23-MAR-2001; 2001WO-US09231.
 23-MAR-2000; 2000US-191637P.
 11-JUL-2000; 2000US-0614150.
 (PEXE) PE CORP NY.
 Venter JC, Adams M, Li PWD, Myers EW;
 WPI; 2001-656860/75.
 N-PSDB; ABL06205.

New isolated nucleic acid detection reagent for detecting 1000 or more genes from Drosophila and for elucidating cell signalling and cell-cell interactions

Disclosure; SEQ ID NO 13098; 21pp + Sequence Listing; English.

The invention relates to an isolated nucleic acid detection reagent capable of detecting 1000 or more genes from Drosophila. The invention is useful in developmental biology and in elucidating cell signalling and cell-cell interactions in higher eukaryotes for the development of insecticides, therapeutics and pharmaceutical drugs. The invention discloses genomic DNA sequences (ABL16176-ABL30511); expressed DNA sequences (ABL01840-ABL16175) and the encoded proteins (ABB57737-ABB72072).

The sequence data for this patent did not form part of the printed specification, but was obtained in electronic format directly from WIPO at ftp.wipo.int/pub/published_pct_sequences.

SQ Sequence 866 AA;

Query Match 1.2%; Score 9; DB 22; Length 866;

Best Local Similarity 100.0%; Pred. No. 8.4;
 Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 615 RDGVGDGQL 623

DB 706 RDGVGDGQL 714

RESULT 7

ABG40429
 ID ABG40429 standard; Peptide; 71 AA.

CC The polypeptide and polynucleotide sequences have applications in
CC diagnostics, forensics, gene mapping, identification of mutations
CC responsible for genetic disorders or other traits to assess biodiversity
CC and to produce other types of data and products dependent on DNA and
CC amino acid sequences. AG00010-ABG30377 represent novel human
CC diagnostic amino acid sequences of the invention.
CC Note: The sequence data for this patent did not appear in the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published_pct_sequences.

XX Sequence 140 AA;

Query Match 1.0%; Score 8; DB 22; Length 140;

Best Local Similarity 100.0%; Pred. No. 17;

Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 748 YAHKLAFL 755

Db 113 YAHKLAFL 120

RESULT 9

AG20040

ID AAG20040 standard; Protein; 254 AA.

XX AC

XX AC

XX AC

XX AC

XX AC

XX AC

XX AC

XX AC

XX AC

XX AC

XX AC

XX AC

XX AC

XX AC

XX AC

XX AC

XX AC

XX AC

XX AC

XX AC

XX AC

XX AC

XX AC

XX AC

XX AC

XX AC

XX AC

XX AC

XX AC

XX AC

XX AC

XX AC

XX AC

XX AC

XX AC

XX AC

XX AC

XX AC

XX AC

XX AC

XX AC

XX AC

XX AC

XX AC

XX AC

XX AC

XX AC

XX AC

XX AC

XX AC

XX AC

XX AC

XX AC

XX AC

XX AC

XX AC

```
PR 02-AUG-1999; 99US-0146388;
PR 02-AUG-1999; 99US-0146389;
PR 03-AUG-1999; 99US-0147038;
PR 04-AUG-1999; 99US-0147204;
PR 04-AUG-1999; 99US-0147302;
PR 05-AUG-1999; 99US-0147192;
PR 05-AUG-1999; 99US-0147260;
PR 06-AUG-1999; 99US-0147303;
PR 06-AUG-1999; 99US-0147416;
PR 09-AUG-1999; 99US-0147493;
PR 09-AUG-1999; 99US-0147935;
PR 10-AUG-1999; 99US-0147935;
PR 11-AUG-1999; 99US-0148171;
PR 11-AUG-1999; 99US-0148319;
PR 12-AUG-1999; 99US-0148341;
PR 13-AUG-1999; 99US-0148565;
PR 13-AUG-1999; 99US-0148684;
PR 16-AUG-1999; 99US-0149368;
PR 17-AUG-1999; 99US-0149175;
PR 18-AUG-1999; 99US-0149426;
PR 20-AUG-1999; 99US-0149722;
PR 20-AUG-1999; 99US-0149723;
PR 20-AUG-1999; 99US-0149929;
PR 23-AUG-1999; 99US-0149902;
PR 23-AUG-1999; 99US-0149930;
PR 25-AUG-1999; 99US-0150566;
PR 26-AUG-1999; 99US-0150884;
PR 27-AUG-1999; 99US-0151065;
PR 27-AUG-1999; 99US-0151066;
PR 27-AUG-1999; 99US-0151080;
PR 30-AUG-1999; 99US-0151303;
PR 31-AUG-1999; 99US-0151438;
PR 01-SEP-1999; 99US-0151930;
PR 07-SEP-1999; 99US-0152363;
PR 10-SEP-1999; 99US-0153070;
PR 13-SEP-1999; 99US-0153758;
PR 15-SEP-1999; 99US-0154018;
PR 16-SEP-1999; 99US-0154039;
PR 20-SEP-1999; 99US-0154779;
PR 22-SEP-1999; 99US-0155139;
PR 23-SEP-1999; 99US-0155486;
PR 24-SEP-1999; 99US-0155659;
PR 28-SEP-1999; 99US-0156458;
PR 29-SEP-1999; 99US-0156596;
PR 04-OCT-1999; 99US-0157117;
PR 05-OCT-1999; 99US-0157253;
PR 06-OCT-1999; 99US-0157865;
PR 07-OCT-1999; 99US-0158029;
PR 08-OCT-1999; 99US-0158232;
PR 12-OCT-1999; 99US-0158369;
PR 13-OCT-1999; 99US-0159293;
PR 13-OCT-1999; 99US-0159294;
PR 13-OCT-1999; 99US-0159295;
PR 14-OCT-1999; 99US-0159330;
PR 14-OCT-1999; 99US-0159331;
PR 14-OCT-1999; 99US-0159637;
PR 14-OCT-1999; 99US-0159638;
PR 16-OCT-1999; 99US-0159584;
PR 21-OCT-1999; 99US-0160741;
PR 21-OCT-1999; 99US-0160767;
PR 21-OCT-1999; 99US-0160768;
PR 21-OCT-1999; 99US-0160770;
PR 21-OCT-1999; 99US-0160814;
PR 21-OCT-1999; 99US-0160815;
PR 22-OCT-1999; 99US-0160980;
PR 22-OCT-1999; 99US-0160981;
PR 22-OCT-1999; 99US-0160989;
PR 25-OCT-1999; 99US-0161404;
PR 25-OCT-1999; 99US-0161405;
PR 25-OCT-1999; 99US-0161406;
PR 26-OCT-1999; 99US-0161359;
PR 26-OCT-1999; 99US-0161360;
PR 26-OCT-1999; 99US-0161361;
PR 26-OCT-1999; 99US-0161361;

PR 28-OCT-1999; 99US-0161920;
PR 28-OCT-1999; 99US-0161992;
PR 28-OCT-1999; 99US-0161993;
PR 29-OCT-1999; 99US-0162142;

Query Match 1.0%; Score 8; DB 21; Length 254;
Best Local Similarity 100.0%; Pred.No. 30;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 702 SVSEPTHYN 709
Db 132 SVSEPTHYN 139

RESULT 10
RAG20039
ID AAG20039 standard; Protein; 271 AA.
XX AC AAG20039;
XX DT 17-OCT-2000 (first entry)
XX DE Arabidopsis thaliana protein fragment SEQ ID NO: 22071.
XX KW Protein identification; signal transduction pathway; metabolic pathway;
KW hybridisation assay; genetic mapping; gene expression control; promoter;
XX KW termination sequence.
XX OS Arabidopsis thaliana.
XX PN EP1033405-A2.
XX PD 06-SEP-2000.
XX PF 25-FEB-2000; 2000EP-0301439.
XX PR 25-FEB-1999; 99US-0121825.
PR 05-MAR-1999; 99US-0123180.
PR 09-MAR-1999; 99US-0123548.
PR 23-MAR-1999; 99US-0125788.
PR 25-MAR-1999; 99US-0126264.
PR 29-MAR-1999; 99US-0126785.
PR 01-APR-1999; 99US-0127462.
PR 06-APR-1999; 99US-0128234.
PR 08-APR-1999; 99US-0128714.
PR 16-APR-1999; 99US-0129845.
PR 19-APR-1999; 99US-0130077.
PR 21-APR-1999; 99US-0130449.
PR 23-APR-1999; 99US-0130510.
PR 23-APR-1999; 99US-0130891.
PR 28-APR-1999; 99US-0131449.
PR 30-APR-1999; 99US-0132048.
PR 30-APR-1999; 99US-0132407.
PR 04-MAY-1999; 99US-0132484.
PR 05-MAY-1999; 99US-0132485.
PR 06-MAY-1999; 99US-0132486.
PR 06-MAY-1999; 99US-0132487.
PR 07-MAY-1999; 99US-0132863.
PR 11-MAY-1999; 99US-0134256.
PR 14-MAY-1999; 99US-0134218.
PR 14-MAY-1999; 99US-0134219.
PR 14-MAY-1999; 99US-0134221.
PR 14-MAY-1999; 99US-0134370.
PR 18-MAY-1999; 99US-0134768.
PR 19-MAY-1999; 99US-0134941.
PR 20-MAY-1999; 99US-0135124.
PR 21-MAY-1999; 99US-0135553.
PR 24-MAY-1999; 99US-0135629.
PR 25-MAY-1999; 99US-0136021.
PR 27-MAY-1999; 99US-0136392.
PR 28-MAY-1999; 99US-0136782.
PR 01-JUN-1999; 99US-0137222.
PR 03-JUN-1999; 99US-0137528.
```

PR 04-JUN-1999; 99US-0137502.
PR 07-JUN-1999; 99US-0137724.
PR 08-JUN-1999; 99US-0138094.
PR 09-JUN-1999; 99US-0138540.
PR 10-JUN-1999; 99US-0138847.
PR 14-JUN-1999; 99US-0139119.
PR 16-JUN-1999; 99US-0139452.
PR 16-JUN-1999; 99US-0139453.
PR 17-JUN-1999; 99US-0139492.
PR 18-JUN-1999; 99US-0139454.
PR 18-JUN-1999; 99US-0139455.
PR 18-JUN-1999; 99US-0139456.
PR 18-JUN-1999; 99US-0139457.
PR 18-JUN-1999; 99US-0139458.
PR 18-JUN-1999; 99US-0139459.
PR 18-JUN-1999; 99US-0139460.
PR 18-JUN-1999; 99US-0139461.
PR 18-JUN-1999; 99US-0139462.
PR 18-JUN-1999; 99US-0139463.
PR 18-JUN-1999; 99US-0139750.
PR 18-JUN-1999; 99US-0139763.
PR 21-JUN-1999; 99US-0139817.
PR 22-JUN-1999; 99US-0139899.
PR 23-JUN-1999; 99US-0140353.
PR 23-JUN-1999; 99US-0140354.
PR 24-JUN-1999; 99US-0140695.
PR 28-JUN-1999; 99US-0140823.
PR 29-JUN-1999; 99US-0140991.
PR 30-JUN-1999; 99US-0141287.
PR 01-JUL-1999; 99US-0141842.
PR 01-JUL-1999; 99US-0142154.
PR 02-JUL-1999; 99US-0142055.
PR 06-JUL-1999; 99US-0142390.
PR 08-JUL-1999; 99US-0142803.
PR 09-JUL-1999; 99US-0142920.
PR 12-JUL-1999; 99US-0143277.
PR 13-JUL-1999; 99US-0143542.
PR 14-JUL-1999; 99US-0143624.
PR 15-JUL-1999; 99US-0144005.
PR 16-JUL-1999; 99US-0144085.
PR 16-JUL-1999; 99US-0144086.
PR 19-JUL-1999; 99US-0144325.
PR 19-JUL-1999; 99US-0144331.
PR 19-JUL-1999; 99US-0144332.
PR 19-JUL-1999; 99US-0144333.
PR 19-JUL-1999; 99US-0144334.
PR 19-JUL-1999; 99US-0144335.
PR 20-JUL-1999; 99US-0144352.
PR 20-JUL-1999; 99US-0144632.
PR 20-JUL-1999; 99US-0144884.
PR 21-JUL-1999; 99US-0144814.
PR 21-JUL-1999; 99US-0145086.
PR 21-JUL-1999; 99US-0145088.
PR 22-JUL-1999; 99US-0145085.
PR 22-JUL-1999; 99US-0145087.
PR 22-JUL-1999; 99US-0145192.
PR 23-JUL-1999; 99US-0145145.
PR 23-JUL-1999; 99US-0145218.
PR 23-JUL-1999; 99US-0145224.
PR 26-JUL-1999; 99US-0145276.
PR 27-JUL-1999; 99US-0145913.
PR 27-JUL-1999; 99US-0145918.
PR 27-JUL-1999; 99US-0145919.
PR 28-JUL-1999; 99US-0145951.
PR 02-AUG-1999; 99US-0146386.
PR 02-AUG-1999; 99US-0146388.
PR 02-AUG-1999; 99US-0146389.
PR 03-AUG-1999; 99US-0147038.
PR 04-AUG-1999; 99US-0147204.
PR 04-AUG-1999; 99US-0147302.
PR 05-AUG-1999; 99US-0147192.
PR 05-AUG-1999; 99US-0147260.

PR 06-AUG-1999; 99US-0147303.
PR 06-AUG-1999; 99US-0147416.
PR 09-AUG-1999; 99US-0147493.
PR 09-AUG-1999; 99US-0147935.
PR 10-AUG-1999; 99US-0148171.
PR 11-AUG-1999; 99US-0148319.
PR 12-AUG-1999; 99US-0148341.
PR 13-AUG-1999; 99US-0148565.
PR 13-AUG-1999; 99US-0148684.
PR 16-AUG-1999; 99US-0149368.
PR 17-AUG-1999; 99US-0149175.
PR 18-AUG-1999; 99US-0149426.
PR 20-AUG-1999; 99US-0149722.
PR 20-AUG-1999; 99US-0149723.
PR 20-AUG-1999; 99US-0149929.
PR 23-AUG-1999; 99US-0149902.
PR 25-AUG-1999; 99US-0149930.
PR 25-AUG-1999; 99US-0150566.
PR 26-AUG-1999; 99US-0150884.
PR 27-AUG-1999; 99US-0151065.
PR 27-AUG-1999; 99US-0151066.
PR 27-AUG-1999; 99US-0151080.
PR 30-AUG-1999; 99US-0151303.
PR 31-AUG-1999; 99US-0151438.
PR 01-SEP-1999; 99US-0151930.
PR 07-SEP-1999; 99US-0152363.
PR 10-SEP-1999; 99US-0153070.
PR 13-SEP-1999; 99US-0153758.
PR 15-SEP-1999; 99US-0154018.
PR 16-SEP-1999; 99US-0154039.
PR 20-SEP-1999; 99US-0154779.
PR 22-SEP-1999; 99US-0155139.
PR 23-SEP-1999; 99US-0155486.
PR 24-SEP-1999; 99US-0155659.
PR 28-SEP-1999; 99US-0156458.
PR 29-SEP-1999; 99US-0156596.
PR 04-OCT-1999; 99US-0157117.
PR 05-OCT-1999; 99US-0157753.
PR 06-OCT-1999; 99US-0157865.
PR 07-OCT-1999; 99US-0158029.
PR 08-OCT-1999; 99US-0158232.
PR 12-OCT-1999; 99US-0158369.
PR 13-OCT-1999; 99US-0159293.
PR 13-OCT-1999; 99US-0159294.
PR 13-OCT-1999; 99US-0159295.
PR 14-OCT-1999; 99US-0159329.
PR 14-OCT-1999; 99US-0159330.
PR 14-OCT-1999; 99US-0159331.
PR 14-OCT-1999; 99US-0159637.
PR 18-OCT-1999; 99US-0159638.
PR 21-OCT-1999; 99US-0160741.
PR 21-OCT-1999; 99US-0160767.
PR 21-OCT-1999; 99US-0160768.
PR 21-OCT-1999; 99US-0160770.
PR 21-OCT-1999; 99US-0160814.
PR 21-OCT-1999; 99US-0160815.
PR 22-OCT-1999; 99US-0160980.
PR 22-OCT-1999; 99US-0160981.
PR 22-OCT-1999; 99US-0160989.
PR 25-OCT-1999; 99US-0161404.
PR 25-OCT-1999; 99US-0161405.
PR 25-OCT-1999; 99US-0161406.
PR 26-OCT-1999; 99US-0161359.
PR 26-OCT-1999; 99US-0161360.
PR 26-OCT-1999; 99US-0161361.
PR 28-OCT-1999; 99US-0161920.
PR 28-OCT-1999; 99US-0161992.
PR 28-OCT-1999; 99US-0161993.
PR 29-OCT-1999; 99US-0162142.

Query Match 1.0%; Score 8; DB 21; Length 271;
Best Local Similarity 100.0%; Pred. No. 31;

	Matches	8;	Conservative	0;	Mismatches	0;	Indels	0;	Gaps	0;
Qy	702 SVSPETHYN 709									
Db	149 SVSPETHYN 156									
RESULT 11										
AAAG20038										
ID	AAAG20038 standard; Protein; 277 AA.									
XX										
AC	AAAG20038;									
XX										
DT	17-OCT-2000 (first entry)									
XX										
DE	Arabidopsis thaliana protein fragment SEQ ID NO: 22070.									
XX										
KW	Protein identification; signal transduction pathway; metabolic pathway;									
KW	hybridisation assay; Genetic mapping; gene expression control; promoter;									
KW	termination sequence.									
XX										
OS	Arabidopsis thaliana.									
XX										
PN	EP1033405-A2.									
XX										
PD	06-SEP-2000.									
XX										
PF	25-FEB-2000; 2000EP-0301439.									
XX										
PR	25-FEB-1999; 99US-0121825.									
PR	05-MAR-1999; 99US-0123180.									
PR	09-MAR-1999; 99US-0123548.									
PR	23-MAR-1999; 99US-0125788.									
PR	25-MAR-1999; 99US-0126264.									
PR	29-MAR-1999; 99US-0126785.									
PR	01-APR-1999; 99US-0127462.									
PR	06-APR-1999; 99US-0128234.									
PR	08-APR-1999; 99US-0128714.									
PR	16-APR-1999; 99US-0129845.									
PR	19-APR-1999; 99US-0130077.									
PR	21-APR-1999; 99US-0130449.									
PR	23-APR-1999; 99US-0130510.									
PR	23-APR-1999; 99US-0130891.									
PR	28-APR-1999; 99US-0131449.									
PR	30-APR-1999; 99US-0132048.									
PR	30-APR-1999; 99US-0132407.									
PR	04-MAY-1999; 99US-0132484.									
PR	05-MAY-1999; 99US-0132485.									
PR	06-MAY-1999; 99US-0132486.									
PR	06-MAY-1999; 99US-0132487.									
PR	07-MAY-1999; 99US-0132863.									
PR	11-MAY-1999; 99US-0134256.									
PR	14-MAY-1999; 99US-0134218.									
PR	14-MAY-1999; 99US-0134219.									
PR	14-MAY-1999; 99US-0134221.									
PR	14-MAY-1999; 99US-0134370.									
PR	18-MAY-1999; 99US-0134768.									
PR	19-MAY-1999; 99US-0134941.									
PR	20-MAY-1999; 99US-0135124.									
PR	21-MAY-1999; 99US-0135353.									
PR	24-MAY-1999; 99US-0135629.									
PR	25-MAY-1999; 99US-0136021.									
PR	27-MAY-1999; 99US-0136392.									

PR 13-AUG-1999; 99US-0148565.
PR 13-AUG-1999; 99US-0148684.
PR 16-AUG-1999; 99US-0149368.
PR 17-AUG-1999; 99US-0149175.
PR 18-AUG-1999; 99US-0149426.
PR 20-AUG-1999; 99US-0149722.
PR 20-AUG-1999; 99US-0149723.
PR 20-AUG-1999; 99US-0149929.
PR 23-AUG-1999; 99US-0149902.
PR 23-AUG-1999; 99US-0149930.
PR 25-AUG-1999; 99US-0149930.
PR 25-AUG-1999; 99US-0150566.
PR 26-AUG-1999; 99US-0150884.
PR 27-AUG-1999; 99US-0151065.
PR 27-AUG-1999; 99US-0151066.
PR 27-AUG-1999; 99US-0151080.
PR 30-AUG-1999; 99US-0151303.
PR 31-AUG-1999; 99US-0151438.
PR 31-AUG-1999; 99US-0151930.
PR 07-SEP-1999; 99US-0152363.
PR 10-SEP-1999; 99US-0153070.
PR 13-SEP-1999; 99US-0153758.
PR 15-SEP-1999; 99US-0154018.
PR 16-SEP-1999; 99US-0154039.
PR 20-SEP-1999; 99US-0154779.
PR 22-SEP-1999; 99US-0155139.
PR 23-SEP-1999; 99US-0155486.
PR 24-SEP-1999; 99US-0155659.
PR 28-SEP-1999; 99US-0156458.
PR 29-SEP-1999; 99US-0156596.
PR 04-OCT-1999; 99US-0157117.
PR 05-OCT-1999; 99US-0157753.
PR 06-OCT-1999; 99US-0157865.
PR 07-OCT-1999; 99US-0158029.
PR 08-OCT-1999; 99US-0158232.
PR 12-OCT-1999; 99US-0158369.
PR 13-OCT-1999; 99US-0159293.
PR 13-OCT-1999; 99US-0159294.
PR 13-OCT-1999; 99US-0159295.
PR 14-OCT-1999; 99US-0159329.
PR 14-OCT-1999; 99US-0159330.
PR 14-OCT-1999; 99US-0159331.
PR 14-OCT-1999; 99US-0159637.
PR 14-OCT-1999; 99US-0159638.
PR 18-OCT-1999; 99US-0159584.
PR 21-OCT-1999; 99US-0160741.
PR 21-OCT-1999; 99US-0160767.
PR 21-OCT-1999; 99US-0160768.
PR 21-OCT-1999; 99US-0160770.
PR 21-OCT-1999; 99US-0160814.
PR 21-OCT-1999; 99US-0160815.
PR 22-OCT-1999; 99US-0160980.
PR 22-OCT-1999; 99US-0160981.
PR 22-OCT-1999; 99US-0160989.
PR 25-OCT-1999; 99US-0161404.
PR 25-OCT-1999; 99US-0161405.
PR 25-OCT-1999; 99US-0161406.
PR 26-OCT-1999; 99US-0161359.
PR 26-OCT-1999; 99US-0161360.
PR 26-OCT-1999; 99US-0161361.
PR 28-OCT-1999; 99US-0161920.
PR 28-OCT-1999; 99US-0161992.
PR 28-OCT-1999; 99US-0161993.
PR 29-OCT-1999; 99US-0162142.

Query Match 8; Score 8; DB 21; Length 277;
Best Local Similarity 100.0%; Pred. No. 32;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 702 SVSPHTYN 709
Db 155 SVSPHTYN 162

RESULT 12
AAG48143
ID AAG48143 standard; Protein; 277 AA.
XX AC AAG48143;
XX DI 18-OCT-2000 (first entry)
XX DE Arabidopsis thaliana protein fragment SEQ ID NO: 60766.
XX KW Protein identification; signal transduction pathway; metabolic pathway;
XX KW hybridisation assay; genetic mapping; gene expression control; promoter;
XX KW termination sequence.
XX OS Arabidopsis thaliana.
XX PN EP1033405-A2.
XX PD 06-SEP-2000.
XX PF 25-FEB-2000; 2000EP-0301439.
XX PR 25-FEB-1999; 99US-0121825.
PR 05-MAR-1999; 99US-0123180.
PR 09-MAR-1999; 99US-0123548.
PR 23-MAR-1999; 99US-0125788.
PR 25-MAR-1999; 99US-0126284.
PR 29-MAR-1999; 99US-0126785.
PR 01-APR-1999; 99US-0127462.
PR 06-APR-1999; 99US-0128234.
PR 08-APR-1999; 99US-0128714.
PR 16-APR-1999; 99US-0129845.
PR 19-APR-1999; 99US-0130077.
PR 21-APR-1999; 99US-0130449.
PR 23-APR-1999; 99US-0130510.
PR 23-APR-1999; 99US-0130891.
PR 28-APR-1999; 99US-0131449.
PR 30-APR-1999; 99US-0132048.
PR 30-APR-1999; 99US-0132407.
PR 04-MAY-1999; 99US-0132484.
PR 05-MAY-1999; 99US-0132485.
PR 06-MAY-1999; 99US-0132486.
PR 06-MAY-1999; 99US-0132487.
PR 07-MAY-1999; 99US-0132863.
PR 11-MAY-1999; 99US-0134256.
PR 14-MAY-1999; 99US-0134218.
PR 14-MAY-1999; 99US-0134219.
PR 14-MAY-1999; 99US-0134221.
PR 14-MAY-1999; 99US-0134370.
PR 18-MAY-1999; 99US-0134768.
PR 19-MAY-1999; 99US-0134941.
PR 20-MAY-1999; 99US-0135124.
PR 21-MAY-1999; 99US-0135353.
PR 24-MAY-1999; 99US-0135629.
PR 25-MAY-1999; 99US-0136021.
PR 27-MAY-1999; 99US-0136392.
PR 28-MAY-1999; 99US-0136782.
PR 01-JUN-1999; 99US-0137222.
PR 03-JUN-1999; 99US-0137528.
PR 04-JUN-1999; 99US-0137502.
PR 07-JUN-1999; 99US-0137724.
PR 08-JUN-1999; 99US-0138094.
PR 10-JUN-1999; 99US-0138540.
PR 10-JUN-1999; 99US-0138847.
PR 14-JUN-1999; 99US-0139119.
PR 16-JUN-1999; 99US-0139452.
PR 16-JUN-1999; 99US-0139453.
PR 17-JUN-1999; 99US-0139492.
PR 18-JUN-1999; 99US-0139454.
PR 18-JUN-1999; 99US-0139455.
PR 18-JUN-1999; 99US-0139456.
PR 18-JUN-1999; 99US-0139457.
PR 18-JUN-1999; 99US-0139458.

PR 18-JUN-1999; 99US-0139459.
PR 18-JUN-1999; 99US-0139460.
PR 18-JUN-1999; 99US-0139461.
PR 18-JUN-1999; 99US-0139462.
PR 18-JUN-1999; 99US-0139463.
PR 18-JUN-1999; 99US-0139750.
PR 18-JUN-1999; 99US-0139763.
PR 21-JUN-1999; 99US-0139817.
PR 22-JUN-1999; 99US-0139899.
PR 23-JUN-1999; 99US-0140353.
PR 23-JUN-1999; 99US-0140354.
PR 24-JUN-1999; 99US-0140695.
PR 28-JUN-1999; 99US-0140823.
PR 29-JUN-1999; 99US-0140991.
PR 30-JUN-1999; 99US-0141287.
PR 01-JUL-1999; 99US-0141842.
PR 01-JUL-1999; 99US-0142154.
PR 02-JUL-1999; 99US-0142055.
PR 06-JUL-1999; 99US-0142390.
PR 08-JUL-1999; 99US-0142803.
PR 09-JUL-1999; 99US-0142920.
PR 12-JUL-1999; 99US-0142977.
PR 13-JUL-1999; 99US-0143542.
PR 14-JUL-1999; 99US-0143624.
PR 15-JUL-1999; 99US-0144005.
PR 16-JUL-1999; 99US-0144085.
PR 19-JUL-1999; 99US-0144086.
PR 19-JUL-1999; 99US-0144331.
PR 19-JUL-1999; 99US-0144332.
PR 19-JUL-1999; 99US-0144333.
PR 19-JUL-1999; 99US-0144334.
PR 19-JUL-1999; 99US-0144335.
PR 20-JUL-1999; 99US-0144352.
PR 20-JUL-1999; 99US-0144632.
PR 20-JUL-1999; 99US-0144884.
PR 21-JUL-1999; 99US-0144814.
PR 21-JUL-1999; 99US-0145086.
PR 21-JUL-1999; 99US-0145088.
PR 22-JUL-1999; 99US-0145085.
PR 22-JUL-1999; 99US-0145087.
PR 22-JUL-1999; 99US-0145089.
PR 22-JUL-1999; 99US-0145192.
PR 23-JUL-1999; 99US-0145145.
PR 23-JUL-1999; 99US-0145218.
PR 23-JUL-1999; 99US-0145224.
PR 26-JUL-1999; 99US-0145276.
PR 27-JUL-1999; 99US-0145913.
PR 27-JUL-1999; 99US-0145918.
PR 27-JUL-1999; 99US-0145919.
PR 28-JUL-1999; 99US-0145951.
PR 02-AUG-1999; 99US-0146386.
PR 02-AUG-1999; 99US-0146388.
PR 02-AUG-1999; 99US-0146389.
PR 03-AUG-1999; 99US-0147038.
PR 04-AUG-1999; 99US-0147204.
PR 05-AUG-1999; 99US-0147302.
PR 05-AUG-1999; 99US-0147192.
PR 06-AUG-1999; 99US-0147260.
PR 06-AUG-1999; 99US-0147303.
PR 06-AUG-1999; 99US-0147416.
PR 09-AUG-1999; 99US-0147493.
PR 09-AUG-1999; 99US-0147935.
PR 10-AUG-1999; 99US-0148171.
PR 11-AUG-1999; 99US-0148319.
PR 12-AUG-1999; 99US-0148341.
PR 13-AUG-1999; 99US-0148565.
PR 13-AUG-1999; 99US-0148684.
PR 16-AUG-1999; 99US-0149368.
PR 17-AUG-1999; 99US-0149175.
PR 18-AUG-1999; 99US-0149426.
PR 20-AUG-1999; 99US-0149722.
PR 20-AUG-1999; 99US-0149723.

PR 20-AUG-1999; 99US-0149929.
PR 23-AUG-1999; 99US-0149902.
PR 23-AUG-1999; 99US-0149930.
PR 25-AUG-1999; 99US-0150566.
PR 26-AUG-1999; 99US-0150884.
PR 27-AUG-1999; 99US-0151065.
PR 27-AUG-1999; 99US-0151066.
PR 27-AUG-1999; 99US-0151080.
PR 30-AUG-1999; 99US-0151303.
PR 31-AUG-1999; 99US-0151438.
PR 01-SEP-1999; 99US-0151930.
PR 07-SEP-1999; 99US-0152363.
PR 10-SEP-1999; 99US-0153070.
PR 13-SEP-1999; 99US-0153758.
PR 15-SEP-1999; 99US-0154018.
PR 16-SEP-1999; 99US-0154039.
PR 20-SEP-1999; 99US-0154779.
PR 22-SEP-1999; 99US-0155139.
PR 23-SEP-1999; 99US-0155486.
PR 24-SEP-1999; 99US-0155659.
PR 28-SEP-1999; 99US-0156458.
PR 29-SEP-1999; 99US-0156596.
PR 04-OCT-1999; 99US-0157117.
PR 05-OCT-1999; 99US-0157753.
PR 06-OCT-1999; 99US-0157865.
PR 07-OCT-1999; 99US-0158029.
PR 08-OCT-1999; 99US-0158232.
PR 12-OCT-1999; 99US-0158369.
PR 13-OCT-1999; 99US-0159293.
PR 13-OCT-1999; 99US-0159294.
PR 13-OCT-1999; 99US-0159295.
PR 14-OCT-1999; 99US-0159329.
PR 14-OCT-1999; 99US-0159330.
PR 14-OCT-1999; 99US-0159331.
PR 14-OCT-1999; 99US-0159637.
PR 14-OCT-1999; 99US-0159638.
PR 18-OCT-1999; 99US-0159584.
PR 21-OCT-1999; 99US-0160741.
PR 21-OCT-1999; 99US-0160767.
PR 21-OCT-1999; 99US-0160768.
PR 21-OCT-1999; 99US-0160770.
PR 21-OCT-1999; 99US-0160814.
PR 22-OCT-1999; 99US-0160815.
PR 22-OCT-1999; 99US-0160980.
PR 22-OCT-1999; 99US-0160981.
PR 22-OCT-1999; 99US-0160989.
PR 25-OCT-1999; 99US-0161404.
PR 25-OCT-1999; 99US-0161405.
PR 25-OCT-1999; 99US-0161406.
PR 26-OCT-1999; 99US-0161359.
PR 26-OCT-1999; 99US-0161360.
PR 26-OCT-1999; 99US-0161361.
PR 28-OCT-1999; 99US-0161920.
PR 28-OCT-1999; 99US-0161992.
PR 28-OCT-1999; 99US-0161993.
PR 29-OCT-1999; 99US-0162142.

Query Match 1.0%; Score 8; DB 21; Length 277;
Best Local Similarity 100.0%; Pred.No. 32;

Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 702 SVSPHYN 709

Db 155 SVSPHYN 162

RESULT 13

AAB92722
ID AAB92722 standard; Protein; 371 AA.

XX AAB92722;

XX AAB92722;

DT 26-JUN-2001 (first entry)

XX DE Human protein sequence SEQ ID NO:11144.
XX KW Human; primer; detection; diagnosis; antisense therapy; gene therapy.
XX OS Homo sapiens.
XX PN EP1074617-A2.
XX PD 07-FEB-2001.
XX PF 28-JUL-2000; 2000EP-0116126.
XX PR 29-JUL-1999; 99JP-0248036.
XX PR 27-AUG-1999; 99JP-0300353.
XX PR 11-JAN-2000; 2000JP-0118776.
XX PR 02-MAY-2000; 2000JP-0183767.
XX PR 09-JUN-2000; 2000JP-0241899.
XX PA (HELI-) HELIX RES INST.
XX PI Ota T, Isogai T, Nishikawa T, Hayashi K, Saito K, Yamamoto J;
XX PI Ishii S, Sugiyama T, Wakamatsu A, Nagai K, Otsuki T;
XX DR WPI; 2001-318749/34.
XX PT Primer sets for synthesising polynucleotides, particularly the 5602
PT full-length cDNAs defined in the specification, and for the detection
PT and/or diagnosis of the abnormality of the proteins encoded by the
PT full-length cDNAs -
XX PS Claim 8; SEQ ID 11144; 2537pp + CD ROM; English.
XX CC The present invention describes primer sets for synthesising 5602
CC full-length cDNAs defined in the specification. Where a primer set
CC comprises: (a) an oligo-dT primer and an oligonucleotide complementary
CC to the complementary strand of a polynucleotide which comprises one of
CC the 5602 nucleotide sequences defined in the specification, where the
CC oligonucleotide comprises at least 15 nucleotides; or (b) a combination
CC of an oligonucleotide comprising a sequence complementary to the
CC complementary strand of a polynucleotide which comprises a 5'-end
CC sequence and an oligonucleotide comprising a sequence complementary to a
CC polynucleotide which comprises a 3'-end sequence, where the
CC oligonucleotide comprises at least 15 nucleotides and the combination of
CC the 5'-end sequence/3'-end sequence is selected from those defined in
CC the specification. The primer sets can be used in antisense therapy and
CC in gene therapy. The primers are useful for synthesising polynucleotides,
CC particularly full-length cDNAs. The primers are also useful for the
CC detection and/or diagnosis of the abnormality of the proteins encoded by
CC the full-length cDNAs. The primers allow obtaining of the full-length
CC cDNAs easily without any specialised methods. AAH03166 to AAH13628 and
CC AAH13633 to AAH13742 represent human cDNA sequences; AAH92446 to
CC AAH95893 represent human amino acid sequences; and AAH13629 to AAH13632
CC represent oligonucleotides, all of which are used in the exemplification
CC of the present invention.
XX SQ Sequence 371 AA;

Query Match 1.0%; Score 8; DB 22; Length 371;
Best Local Similarity 100.0%; Pred. No. 42;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 748 YAHKLAFL 755
Db 344 YAHKLAFL 351
|||||

RESULT 14
AAG48142
ID AAG48142 standard; Protein; 398 AA.
XX
AC AAG48142;
XX

DT 18-OCT-2000 (first entry)
XX DE Arabidopsis thaliana protein fragment SEQ ID NO: 60765.
XX KW Protein identification; signal transduction pathway; metabolic pathway;
KW hybridisation assay; genetic mapping; gene expression control; promoter;
KW termination sequence.
XX OS Arabidopsis thaliana.
XX PN EP1033405-A2.
XX PD 06-SEP-2000.
XX PF 25-FEB-2000; 2000EP-0301439.
XX PR 25-FEB-1999; 99US-0121825.
XX PR 05-MAR-1999; 99US-0123180.
XX PR 09-MAR-1999; 99US-0123548.
XX PR 23-MAR-1999; 99US-0125788.
XX PR 25-MAR-1999; 99US-0126264.
XX PR 29-MAR-1999; 99US-0126785.
XX PR 01-APR-1999; 99US-0127462.
XX PR 06-APR-1999; 99US-0128234.
XX PR 08-APR-1999; 99US-0128714.
XX PR 16-APR-1999; 99US-0129845.
XX PR 19-APR-1999; 99US-0130077.
XX PR 21-APR-1999; 99US-0130449.
XX PR 23-APR-1999; 99US-0130510.
XX PR 23-APR-1999; 99US-0130891.
XX PR 28-APR-1999; 99US-0131449.
XX PR 30-APR-1999; 99US-0132048.
XX PR 30-APR-1999; 99US-0132407.
XX PR 04-MAY-1999; 99US-0132484.
XX PR 05-MAY-1999; 99US-0132485.
XX PR 06-MAY-1999; 99US-0132486.
XX PR 06-MAY-1999; 99US-0132487.
XX PR 07-MAY-1999; 99US-0132863.
XX PR 11-MAY-1999; 99US-0134256.
XX PR 14-MAY-1999; 99US-0134218.
XX PR 14-MAY-1999; 99US-0134219.
XX PR 14-MAY-1999; 99US-0134221.
XX PR 18-MAY-1999; 99US-0134370.
XX PR 19-MAY-1999; 99US-0134768.
XX PR 20-MAY-1999; 99US-0135124.
XX PR 21-MAY-1999; 99US-0135353.
XX PR 24-MAY-1999; 99US-0135629.
XX PR 25-MAY-1999; 99US-0136021.
XX PR 27-MAY-1999; 99US-0136392.
XX PR 28-MAY-1999; 99US-0136782.
XX PR 01-JUN-1999; 99US-0137222.
XX PR 03-JUN-1999; 99US-0137528.
XX PR 04-JUN-1999; 99US-0137502.
XX PR 07-JUN-1999; 99US-0137724.
XX PR 08-JUN-1999; 99US-0138094.
XX PR 10-JUN-1999; 99US-0138540.
XX PR 10-JUN-1999; 99US-0138847.
XX PR 14-JUN-1999; 99US-0139119.
XX PR 16-JUN-1999; 99US-0139452.
XX PR 16-JUN-1999; 99US-0139453.
XX PR 17-JUN-1999; 99US-0139492.
XX PR 18-JUN-1999; 99US-0139454.
XX PR 18-JUN-1999; 99US-0139455.
XX PR 18-JUN-1999; 99US-0139456.
XX PR 18-JUN-1999; 99US-0139457.
XX PR 18-JUN-1999; 99US-0139458.
XX PR 18-JUN-1999; 99US-0139459.
XX PR 18-JUN-1999; 99US-0139460.
XX PR 18-JUN-1999; 99US-0139461.
XX PR 18-JUN-1999; 99US-0139462.
XX PR 18-JUN-1999; 99US-0139463.
XX PR 18-JUN-1999; 99US-0139750.


```
PR 18-JUN-1999; 99US-0139763.
PR 21-JUN-1999; 99US-0139817.
PR 22-JUN-1999; 99US-0139899.
PR 23-JUN-1999; 99US-0140353.
PR 23-JUN-1999; 99US-0140354.
PR 24-JUN-1999; 99US-0140695.
PR 28-JUN-1999; 99US-0140823.
PR 29-JUN-1999; 99US-0140991.
PR 30-JUN-1999; 99US-0141287.
PR 01-JUL-1999; 99US-0141842.
PR 01-JUL-1999; 99US-0142154.
PR 02-JUL-1999; 99US-0142055.
PR 06-JUL-1999; 99US-0142390.
PR 08-JUL-1999; 99US-0142803.
PR 09-JUL-1999; 99US-0142920.
PR 12-JUL-1999; 99US-0142977.
PR 13-JUL-1999; 99US-0143542.
PR 14-JUL-1999; 99US-0143624.
PR 15-JUL-1999; 99US-0144005.
PR 16-JUL-1999; 99US-0144085.
PR 16-JUL-1999; 99US-0144086.
PR 19-JUL-1999; 99US-0144325.
PR 19-JUL-1999; 99US-0144331.
PR 19-JUL-1999; 99US-0144332.
PR 19-JUL-1999; 99US-0144333.
PR 19-JUL-1999; 99US-0144334.
PR 20-JUL-1999; 99US-0144335.
PR 20-JUL-1999; 99US-0144352.
PR 20-JUL-1999; 99US-0144632.
PR 20-JUL-1999; 99US-0144684.
PR 21-JUL-1999; 99US-0144814.
PR 21-JUL-1999; 99US-0145086.
PR 21-JUL-1999; 99US-0145088.
PR 22-JUL-1999; 99US-0145085.
PR 22-JUL-1999; 99US-0145087.
PR 22-JUL-1999; 99US-0145089.
PR 22-JUL-1999; 99US-0145192.
PR 23-JUL-1999; 99US-0145145.
PR 23-JUL-1999; 99US-0145218.
PR 23-JUL-1999; 99US-0145224.
PR 26-JUL-1999; 99US-0145276.
PR 27-JUL-1999; 99US-0145913.
PR 27-JUL-1999; 99US-0145918.
PR 27-JUL-1999; 99US-0145919.
PR 28-JUL-1999; 99US-0145951.
PR 02-AUG-1999; 99US-0145886.
PR 02-AUG-1999; 99US-0146388.
PR 03-AUG-1999; 99US-0147038.
PR 04-AUG-1999; 99US-0147204.
PR 04-AUG-1999; 99US-0147302.
PR 05-AUG-1999; 99US-0147192.
PR 05-AUG-1999; 99US-0147260.
PR 06-AUG-1999; 99US-0147303.
PR 06-AUG-1999; 99US-0147416.
PR 09-AUG-1999; 99US-0147493.
PR 09-AUG-1999; 99US-0147935.
PR 10-AUG-1999; 99US-0148171.
PR 11-AUG-1999; 99US-0148319.
PR 12-AUG-1999; 99US-0148341.
PR 13-AUG-1999; 99US-0148565.
PR 13-AUG-1999; 99US-0148684.
PR 16-AUG-1999; 99US-0149368.
PR 17-AUG-1999; 99US-0149175.
PR 18-AUG-1999; 99US-0149426.
PR 20-AUG-1999; 99US-0149722.
PR 20-AUG-1999; 99US-0149723.
PR 20-AUG-1999; 99US-0149929.
PR 23-AUG-1999; 99US-0149802.
PR 23-AUG-1999; 99US-0149930.
PR 25-AUG-1999; 99US-0150566.
PR 26-AUG-1999; 99US-0150884.
PR 27-AUG-1999; 99US-0151065.
PR 27-AUG-1999; 99US-0151065.
PR 27-AUG-1999; 99US-0151066.
PR 27-AUG-1999; 99US-0151080.
PR 30-AUG-1999; 99US-0151303.
PR 31-AUG-1999; 99US-0151438.
PR 01-SEP-1999; 99US-0151930.
PR 07-SEP-1999; 99US-0152363.
PR 10-SEP-1999; 99US-0153070.
PR 13-SEP-1999; 99US-0153758.
PR 15-SEP-1999; 99US-0154018.
PR 16-SEP-1999; 99US-0154039.
PR 20-SEP-1999; 99US-0154779.
PR 22-SEP-1999; 99US-0155139.
PR 23-SEP-1999; 99US-0155486.
PR 24-SEP-1999; 99US-0155659.
PR 28-SEP-1999; 99US-0156458.
PR 29-SEP-1999; 99US-0156596.
PR 04-OCT-1999; 99US-0157117.
PR 05-OCT-1999; 99US-0157753.
PR 06-OCT-1999; 99US-0157865.
PR 07-OCT-1999; 99US-0158029.
PR 08-OCT-1999; 99US-0158232.
PR 12-OCT-1999; 99US-0158369.
PR 13-OCT-1999; 99US-0159284.
PR 13-OCT-1999; 99US-0159285.
PR 14-OCT-1999; 99US-0159329.
PR 14-OCT-1999; 99US-0159330.
PR 14-OCT-1999; 99US-0159331.
PR 14-OCT-1999; 99US-0159637.
PR 14-OCT-1999; 99US-0159638.
PR 18-OCT-1999; 99US-0159584.
PR 21-OCT-1999; 99US-0160741.
PR 21-OCT-1999; 99US-0160767.
PR 21-OCT-1999; 99US-0160768.
PR 21-OCT-1999; 99US-0160770.
PR 21-OCT-1999; 99US-0160814.
PR 21-OCT-1999; 99US-0160815.
PR 22-OCT-1999; 99US-0160980.
PR 22-OCT-1999; 99US-0160981.
PR 22-OCT-1999; 99US-0160989.
PR 25-OCT-1999; 99US-0161404.
PR 25-OCT-1999; 99US-0161405.
PR 25-OCT-1999; 99US-0161406.
PR 26-OCT-1999; 99US-0161359.
PR 26-OCT-1999; 99US-0161360.
PR 26-OCT-1999; 99US-0161361.
PR 28-OCT-1999; 99US-0161920.
PR 28-OCT-1999; 99US-0161992.
PR 28-OCT-1999; 99US-0161993.
PR 29-OCT-1999; 99US-0162142.
PR 29-OCT-1999; 99US-0162143.

Query Match 1.0%; Score 8; DB 21; Length 398;
Best Local Similarity 100.0%; Pred.No. 45;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 702 SVSPHYN 709
Db 276 SVSPHYN 283

RESULT 15
AAG48141
ID AAG48141 standard; Protein; 433 AA.
XX
AC AAG48141;
XX
DT 18-OCT-2000 (first entry)
DE Arabidopsis thaliana protein fragment SEQ ID NO: 60764.
KW Protein identification; signal transduction pathway; metabolic pathway;
KW hybridisation assay; genetic mapping; gene expression control; promoter;
KW termination sequence.
```

XX Arabidopsis thaliana.
XX EP1033405-A2.
XX PD 06-SEP-2000.
XX PF 25-FEB-2000; 2000EP-0301439.
XX PR 25-FEB-1999; 99US-0121825.
XX PR 05-MAR-1999; 99US-0123180.
XX PR 09-MAR-1999; 99US-0123548.
XX PR 23-MAR-1999; 99US-0125788.
XX PR 25-MAR-1999; 99US-0126264.
XX PR 29-MAR-1999; 99US-0126785.
XX PR 01-APR-1999; 99US-0127462.
XX PR 06-APR-1999; 99US-0128234.
XX PR 08-APR-1999; 99US-0128714.
XX PR 16-APR-1999; 99US-0129845.
XX PR 19-APR-1999; 99US-0130077.
XX PR 21-APR-1999; 99US-0130449.
XX PR 23-APR-1999; 99US-0130510.
XX PR 28-APR-1999; 99US-0130891.
XX PR 30-APR-1999; 99US-0131449.
XX PR 30-APR-1999; 99US-0132048.
XX PR 04-MAY-1999; 99US-0132484.
XX PR 05-MAY-1999; 99US-0132485.
XX PR 06-MAY-1999; 99US-0132486.
XX PR 07-MAY-1999; 99US-0132863.
XX PR 11-MAY-1999; 99US-0134256.
XX PR 14-MAY-1999; 99US-0134218.
XX PR 14-MAY-1999; 99US-0134219.
XX PR 14-MAY-1999; 99US-0134221.
XX PR 14-MAY-1999; 99US-0134370.
XX PR 18-MAY-1999; 99US-0134768.
XX PR 19-MAY-1999; 99US-0134941.
XX PR 20-MAY-1999; 99US-0135124.
XX PR 21-MAY-1999; 99US-0135353.
XX PR 24-MAY-1999; 99US-0135629.
XX PR 25-MAY-1999; 99US-0136021.
XX PR 27-MAY-1999; 99US-0136392.
XX PR 28-MAY-1999; 99US-0136782.
XX PR 01-JUN-1999; 99US-0137222.
XX PR 03-JUN-1999; 99US-0137528.
XX PR 04-JUN-1999; 99US-0137502.
XX PR 07-JUN-1999; 99US-0137724.
XX PR 08-JUN-1999; 99US-0138094.
XX PR 10-JUN-1999; 99US-0138540.
XX PR 14-JUN-1999; 99US-0139119.
XX PR 16-JUN-1999; 99US-0139452.
XX PR 16-JUN-1999; 99US-0139453.
XX PR 17-JUN-1999; 99US-0139492.
XX PR 18-JUN-1999; 99US-0139454.
XX PR 18-JUN-1999; 99US-0139455.
XX PR 18-JUN-1999; 99US-0139456.
XX PR 18-JUN-1999; 99US-0139457.
XX PR 18-JUN-1999; 99US-0139458.
XX PR 18-JUN-1999; 99US-0139459.
XX PR 18-JUN-1999; 99US-0139460.
XX PR 18-JUN-1999; 99US-0139461.
XX PR 18-JUN-1999; 99US-0139462.
XX PR 18-JUN-1999; 99US-0139463.
XX PR 18-JUN-1999; 99US-0139750.
XX PR 18-JUN-1999; 99US-0139763.
XX PR 21-JUN-1999; 99US-0139817.
XX PR 22-JUN-1999; 99US-0139899.
XX PR 23-JUN-1999; 99US-0140353.
XX PR 23-JUN-1999; 99US-0140354.
XX PR 24-JUN-1999; 99US-0140695.
XX PR 28-JUN-1999; 99US-0140823.
XX PR 29-JUN-1999; 99US-0140991.
XX PR 30-JUN-1999; 99US-0141287.
XX PR 01-JUL-1999; 99US-0141842.
XX PR 01-JUL-1999; 99US-0142154.
XX PR 02-JUL-1999; 99US-0142055.
XX PR 06-JUL-1999; 99US-0142390.
XX PR 08-JUL-1999; 99US-0142803.
XX PR 09-JUL-1999; 99US-0142920.
XX PR 12-JUL-1999; 99US-0142977.
XX PR 13-JUL-1999; 99US-0143542.
XX PR 14-JUL-1999; 99US-0143624.
XX PR 15-JUL-1999; 99US-0144005.
XX PR 16-JUL-1999; 99US-0144085.
XX PR 16-JUL-1999; 99US-0144086.
XX PR 19-JUL-1999; 99US-0144325.
XX PR 19-JUL-1999; 99US-0144331.
XX PR 19-JUL-1999; 99US-0144332.
XX PR 19-JUL-1999; 99US-0144333.
XX PR 19-JUL-1999; 99US-0144334.
XX PR 19-JUL-1999; 99US-0144335.
XX PR 20-JUL-1999; 99US-0144352.
XX PR 20-JUL-1999; 99US-0144632.
XX PR 20-JUL-1999; 99US-0144884.
XX PR 21-JUL-1999; 99US-0144814.
XX PR 21-JUL-1999; 99US-0145086.
XX PR 21-JUL-1999; 99US-0145088.
XX PR 22-JUL-1999; 99US-0145085.
XX PR 22-JUL-1999; 99US-0145087.
XX PR 22-JUL-1999; 99US-0145089.
XX PR 22-JUL-1999; 99US-0145192.
XX PR 23-JUL-1999; 99US-0145145.
XX PR 23-JUL-1999; 99US-0145218.
XX PR 23-JUL-1999; 99US-0145224.
XX PR 26-JUL-1999; 99US-0145276.
XX PR 27-JUL-1999; 99US-0145913.
XX PR 27-JUL-1999; 99US-0145918.
XX PR 27-JUL-1999; 99US-0145919.
XX PR 28-JUL-1999; 99US-0145951.
XX PR 02-AUG-1999; 99US-0146386.
XX PR 02-AUG-1999; 99US-0146388.
XX PR 03-AUG-1999; 99US-0146389.
XX PR 03-AUG-1999; 99US-0147038.
XX PR 04-AUG-1999; 99US-0147204.
XX PR 04-AUG-1999; 99US-0147302.
XX PR 05-AUG-1999; 99US-0147192.
XX PR 05-AUG-1999; 99US-0147260.
XX PR 06-AUG-1999; 99US-0147303.
XX PR 06-AUG-1999; 99US-0147416.
XX PR 09-AUG-1999; 99US-0147493.
XX PR 09-AUG-1999; 99US-0147935.
XX PR 10-AUG-1999; 99US-0148171.
XX PR 11-AUG-1999; 99US-0148319.
XX PR 12-AUG-1999; 99US-0148341.
XX PR 13-AUG-1999; 99US-0148565.
XX PR 13-AUG-1999; 99US-0148684.
XX PR 16-AUG-1999; 99US-0149368.
XX PR 17-AUG-1999; 99US-0149175.
XX PR 18-AUG-1999; 99US-0149426.
XX PR 20-AUG-1999; 99US-0149722.
XX PR 20-AUG-1999; 99US-0149723.
XX PR 20-AUG-1999; 99US-0149929.
XX PR 23-AUG-1999; 99US-0149902.
XX PR 23-AUG-1999; 99US-0149930.
XX PR 25-AUG-1999; 99US-0150566.
XX PR 26-AUG-1999; 99US-0150884.
XX PR 27-AUG-1999; 99US-0151065.
XX PR 27-AUG-1999; 99US-0151066.
XX PR 27-AUG-1999; 99US-0151080.
XX PR 30-AUG-1999; 99US-0151303.
XX PR 31-AUG-1999; 99US-0151438.
XX PR 01-SEP-1999; 99US-0151930.
XX PR 07-SEP-1999; 99US-0152363.
XX PR 10-SEP-1999; 99US-0153070.

```
PR 13-SEP-1999; 99US-0153758.
PR 15-SEP-1999; 99US-0154018.
PR 16-SEP-1999; 99US-0154039.
PR 20-SEP-1999; 99US-0154779.
PR 22-SEP-1999; 99US-0155139.
PR 23-SEP-1999; 99US-0155486.
PR 24-SEP-1999; 99US-0155659.
PR 28-SEP-1999; 99US-0156458.
PR 29-SEP-1999; 99US-0156596.
PR 04-OCT-1999; 99US-0157117.
PR 05-OCT-1999; 99US-0157753.
PR 06-OCT-1999; 99US-0157865.
PR 07-OCT-1999; 99US-0158029.
PR 08-OCT-1999; 99US-0158232.
PR 12-OCT-1999; 99US-0158369.
PR 13-OCT-1999; 99US-0159293.
PR 13-OCT-1999; 99US-0159294.
PR 13-OCT-1999; 99US-0159295.
PR 14-OCT-1999; 99US-0159329.
PR 14-OCT-1999; 99US-0159330.
PR 14-OCT-1999; 99US-0159331.
PR 14-OCT-1999; 99US-0159637.
PR 14-OCT-1999; 99US-0159638.
PR 18-OCT-1999; 99US-0159584.
PR 21-OCT-1999; 99US-0160741.
PR 21-OCT-1999; 99US-0160767.
PR 21-OCT-1999; 99US-0160768.
PR 21-OCT-1999; 99US-0160770.
PR 21-OCT-1999; 99US-0160814.
PR 21-OCT-1999; 99US-0160815.
PR 22-OCT-1999; 99US-0160980.
PR 22-OCT-1999; 99US-0160981.
PR 22-OCT-1999; 99US-0160989.
PR 25-OCT-1999; 99US-0161404.
PR 25-OCT-1999; 99US-0161405.
PR 25-OCT-1999; 99US-0161406.
PR 26-OCT-1999; 99US-0161359.
PR 26-OCT-1999; 99US-0161360.
PR 26-OCT-1999; 99US-0161361.
PR 28-OCT-1999; 99US-0161920.
PR 28-OCT-1999; 99US-0161992.
PR 28-OCT-1999; 99US-0161993.
PR 29-OCT-1999; 99US-0162142.

Query Match 1.0%; Score 8; DB 21; Length 433;
Best Local Similarity 100.0%; Pred. No. 48;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

```
Qy 702 SVSPHYN 709
Db 311 SVSPHYN 318
```

Search completed: December 10, 2003, 17:34:55
Job time : 49 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: December 4, 2003, 19:11:53 ; Search time 41 Seconds
(without alignments)
4877.824 Million cell updates/sec

Title: US-10-043-774B-2

Perfect score: 4102

Sequence: 1 MIFGVNTRQNLDHVKESTG.....VQSIHREPNLSNRLYYL 775

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 830525 seqs, 258052604 residues

Total number of hits satisfying chosen parameters: 830525

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

SPTREMBL 23:*

- 1: sp_archaea:*
- 2: sp_bacteria:*
- 3: sp_fungi:*
- 4: sp_human:*
- 5: sp_invertebrate:*
- 6: sp_mammal:*
- 7: sp_mhc:*
- 8: sp_organelle:*
- 9: sp_phage:*
- 10: sp_plant:*
- 11: sp_rodent:*
- 12: sp_virus:*
- 13: sp_vertebrate:*
- 14: sp_unclassified:*
- 15: sp_virus:*
- 16: sp_bacteriap:*
- 17: sp_archaeap:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	ID	Description
1	4102	100.0	775	4	Q96JD5
2	4087	99.6	861	4	Q96J94
3	4072	99.3	861	4	Q8TBV5
4	4071	99.2	861	4	Q95404
5	3980	97.0	862	11	Q9JMB7
6	3883	94.7	829	4	Q8NA60
7	2828.5	69.0	858	13	Q8UVX0
8	2103.5	51.3	852	4	Q8NEH2
9	2101.5	51.2	852	4	Q8NPV8
10	2009.5	49.0	854	5	Q9GPA8
11	1859.5	45.3	878	11	Q8CCT6
12	1838.5	44.8	866	4	Q8N8G9
13	1811.5	44.2	808	5	Q9GPA7
14	1766	43.1	971	11	Q9JMB6
15	1764	43.0	971	11	Q8CDG1
16	1754	42.8	973	4	Q8TC59

17	1468	35.8	866	5	Q76922	076922 drosophila
18	1428.5	34.8	843	5	Q9VKM1	Q9VKM1 drosophila
19	1389.5	33.9	580	11	Q99MV6	Q99MV6 mus musculus
20	1295.5	31.6	824	5	P90786	P90786 caenorhabdi
21	1230	30.0	530	4	Q96SW6	Q96SW6 homo sapien
22	1209	29.5	421	5	Q9GPA6	Q9GPA6 strongyloce
23	1162	28.3	722	5	Q17567	Q17567 caenorhabdi
24	989	24.1	791	5	Q8ISG8	Q8ISG8 stylonychia
25	962	23.5	371	4	Q9NW28	Q9NW28 homo sapien
26	804.5	19.6	780	5	Q8MXZ9	Q8MXZ9 tetrahymena
27	803	19.6	779	5	Q8MQL1	Q8MQL1 tetrahymena
28	728	17.7	781	5	Q9USC9	Q9USC9 paramedius
29	520.5	12.7	1194	10	Q8SHE2	Q8SHE2 arabidopsis
30	508.5	12.4	860	11	Q8CUG0	Q8CUG0 mus musculus
31	508.5	12.4	1014	10	Q9SHF3	Q9SHF3 arabidopsis
32	507	12.4	978	10	Q8LP00	Q8LP00 oryza sativ
33	505.5	12.3	910	5	O16720	O16720 caenorhabdi
34	504.5	12.3	861	11	Q8CJF8	Q8CJF8 mus musculus
35	502.5	12.3	951	11	Q8BTF4	Q8BTF4 mus musculus
36	502	12.2	860	11	Q8CJF9	Q8CJF9 mus musculus
37	500	12.2	950	5	Q9V3L2	Q9V3L2 drosophila
38	500	12.2	984	5	Q9V6V6	Q9V6V6 drosophila
39	499.5	12.2	857	11	Q8CJG1	Q8CJG1 mus musculus
40	498.5	12.2	1002	5	Q20578	Q20578 caenorhabdi
41	497.5	12.1	924	10	Q9ZVD5	Q9ZVD5 arabidopsis
42	497	12.1	860	4	Q9H9G7	Q9H9G7 homo sapien
43	495.5	12.1	924	4	Q9HCK5	Q9HCK5 homo sapien
44	495	12.1	848	11	Q8CGU0	Q8CGU0 mus musculus
45	491.5	12.0	1123	10	Q9LP83	Q9LP83 arabidopsis

ALIGNMENTS

RESULT 1

Q96JD5	PRELIMINARY;	PRT;	775 AA.
AC	Q96JD5;		
DT	01-DEC-2001 (Tremblrel. 19, Created)		
DT	01-DEC-2001 (Tremblrel. 19, Last sequence update)		
DT	01-OCT-2002 (Tremblrel. 22, Last annotation update)		
DE	HIWI.		
OS	Homo sapiens (Human)		
OC	Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;		
OC	Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.		
OX	NCBI_TaxID=9606;		
RN	[1]		
RP	SEQUENCE FROM N.A.		
RC	TISSUE=Testis;		
RA	Sharma A.K., Nelson M.C., Brandt J.E., Wessman M., Muhud N.,		
RA	Weller K.P., Hoffman R.,		
RT	"Human CD34+ Stem Cells Express the hiwi Gene, a Human Homolog of the		
RT	Drosophila Gene piwi.";		
RL	Submitted (MAY-2000) to the EMBL/GenBank/DBJ databases.		
DR	EMBL; AF264004; AAK92281.1; -		
DR	InterPro; IPR003100; PAZ.		
DR	InterPro; IPR003165; Pwi.		
DR	Pfam; PF02170; PAZ; 1.		
DR	Pfam; PF02171; Pwi; 1.		
DR	PROSITE; PS0821; PAZ; 1.		
DR	PROSITE; PS0822; Pwi; 1.		
SQ	SEQUENCE 775 AA; 89484 NW; DFI69A2E9EAFD916 CRC64;		

Query Match	100.0%;	Score 4102;	DB 4;	Length 775;
Best Local Similarity	100.0%;	Pred. No. 4.1e-318;		
Matches 775;	Conservative 0;	Mismatches 0;	Indels 0;	Gaps 0;
QY	1	MIFGVNTRQNLDHVKESTGSGIIVRLTSRHNHFRITSPQWALYQYHIDYNPLMEARRLR	60	
Db	1	MIFGVNTRQNLDHVKESTGSGIIVRLTSRHNHFRITSPQWALYQYHIDYNPLMEARRLR	60	
QY	61	SALLFOHEDLCKCHAFDGTILFLPKLOOKVTFVSTRNGEDVRIITLTNEIPPTSP	120	

```
DD 61 SALLFQHEDLGKCHAFDGTILFLPRLOQKVTEVFSKTRNGEDVRITITITLNEIPPTSP 120
QY 121 TGLQFYNIIFRLLKIMNLQOQGRNYNPNNDIPSHRLVIWPGFTTSILOYENSIMLC 180
DB 121 TGLQFYNIIFRLLKIMNLQOQGRNYNPNNDIPSHRLVIWPGFTTSILOYENSIMLC 180
QY 181 TDVSHKVLRSSETVLDPMFNFYHQTEBHKFQEQVSKELIGLVLTLYKNNKYRVDDIDWDQ 240
DB 181 TDVSHKVLRSSETVLDPMFNFYHQTEBHKFQEQVSKELIGLVLTLYKNNKYRVDDIDWDQ 240
QY 241 NPKSFVKADGSEVSLEYRYKQYNOEITDLKOPVLVSQPKRRGPGGTLPGPAMLIPEL 300
DB 241 NPKSFVKADGSEVSLEYRYKQYNOEITDLKOPVLVSQPKRRGPGGTLPGPAMLIPEL 300
QY 301 CYLTGLTDRNRDNFNMKDLAVHTRLTTPQOREVGRLLIDYIHKNDNVORELRDGLSPD 360
DB 301 CYLTGLTDRNRDNFNMKDLAVHTRLTTPQOREVGRLLIDYIHKNDNVORELRDGLSPD 360
QY 361 SNLLSFSGRILQTEKIHQGGKTFDYNPQFADWSKETRGAPLISVKPLDNWLLIYTRNVE 420
DB 361 SNLLSFSGRILQTEKIHQGGKTFDYNPQFADWSKETRGAPLISVKPLDNWLLIYTRNVE 420
QY 421 AANSLIQNLFKVTPAMGQMRKAIMLEVDDRTYALVLOQKVTDATQIVVCLLSNRKD 480
DB 421 AANSLIQNLFKVTPAMGQMRKAIMLEVDDRTYALVLOQKVTDATQIVVCLLSNRKD 480
QY 481 KYDAIKKYLCTDPTSCQCVARTLGKQOTVMAIATKIALQMNCKMGELWVDIPLKLV 540
DB 481 KYDAIKKYLCTDPTSCQCVARTLGKQOTVMAIATKIALQMNCKMGELWVDIPLKLV 540
QY 541 MIVGIDCYHDMTAGRSIAGFVASINEGTRWFSRCIFQDRGOELVDGLKVCLOAALRAW 600
DB 541 MIVGIDCYHDMTAGRSIAGFVASINEGTRWFSRCIFQDRGOELVDGLKVCLOAALRAW 600
QY 601 NSCNEVMPRIIYVRDVGQGLKTLVNYEVPQFLDCLKISGRGNPRLTVIVVKRVNT 660
DB 601 NSCNEVMPRIIYVRDVGQGLKTLVNYEVPQFLDCLKISGRGNPRLTVIVVKRVNT 660
QY 661 RFFAQSGRLQNPFGTVIDVEVTRPEWYDFIVSQAVRSVSPHYNVIYDNSGLKPD 720
DB 661 RFFAQSGRLQNPFGTVIDVEVTRPEWYDFIVSQAVRSVSPHYNVIYDNSGLKPD 720
QY 721 HIQRUTYKLCHIIYNNWPGVIRVPAPQYAHKLAFLVGQSIHREPNSLSNRLYYL 775
DB 721 HIQRUTYKLCHIIYNNWPGVIRVPAPQYAHKLAFLVGQSIHREPNSLSNRLYYL 775

RESULT 2
Q96J94 ID Q96J94 PRELIMINARY; PRT; 861 AA.
AC Q96J94;
DT 01-DEC-2001 (TrEMBLrel. 19, Created)
DT 01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
DE PIWI protein.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]_
RP SEQUENCE FROM N.A.
RC TISSUE=Testis;
RA Sha J.H.;
RT "Cloning and identification of human piwi protein related to testis
RL development."
RL Submitted (MAY-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF387507; AAK69348.1; -.
DR InterPro; IPR003100; PAZ.
DR InterPro; IPR003165; Piwi.
DR Pfam; PF02170; PAZ; 1.
DR Pfam; PF02171; Piwi; 1.
DR PROSITE; PS50821; PAZ; 1.
DR PROSITE; PS50822; PIWI; 1.

SQ SEQUENCE 861 AA; 98603 MW; 58D7F6C7321DEFA4 CRC64;
Query Match
Best Local Similarity 99.6%; Score 4087; DB 4; Length 861;
Matches 772; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 4 GVNTQRNLDRHVKESKTSSGIIIVRLSTNNHFLTRSPQWALYQYHIDYNPLMEARLSAL 63
DB 90 GVNTQRNLDRHVKESKTSSGIIIVRLSTNNHFLTRSPQWALYQYHIDYNPLMEARLSAL 149
QY 64 LFOHEDLICKCHAFDGTILFLPKLOQKVTEVFSKTRNGEDVRITITITLNEIPPTSPCL 123
DB 150 LFOHEDLICKCHAFDGTILFLPKLOQKVTEVFSKTRNGEDVRITITITLNEIPPTSPCL 209
QY 124 QFYNIIFRLLKIMNLQOQGRNYNPNNDIPSHRLVIWPGFTTSILOYENSIMLCIDV 183
DB 210 QFYNIIFRLLKIMNLQOQGRNYNPNNDIPSHRLVIWPGFTTSILOYENSIMLCIDV 269
QY 184 SHKVLRSSTVLDPMFNFYHQTEBHKFQEQVSKELIGLVLTLYKNNKYRVDDIDWDQPK 243
DB 270 SHKVLRSSTVLDPMFNFYHQTEBHKFQEQVSKELIGLVLTLYKNNKYRVDDIDWDQPK 329
QY 244 STPKKADGSEVSLEYRYKQYNOEITDLKOPVLVSQPKRRGPGGTLPGPAMLIPEL 303
DB 330 STPKKADGSEVSLEYRYKQYNOEITDLKOPVLVSQPKRRGPGGTLPGPAMLIPEL 389
QY 304 TGLTKMRNDNFMKDLAVHTRLTTPQOREVGRLLIDYIHKNDNVORELRDGLSPD 363
DB 390 TGLTKMRNDNFMKDLAVHTRLTTPQOREVGRLLIDYIHKNDNVORELRDGLSPD 449
QY 364 LSPSGRILQTEKIHQGGKTFDYNPQFADWSKETRGAPLISVKPLDNWLLIYTRNVEAAN 423
DB 450 LSPSGRILQTEKIHQGGKTFDYNPQFADWSKETRGAPLISVKPLDNWLLIYTRNVEAAN 509
QY 424 SLIQNLFKVTPAMGQMRKAIMLEVDDRTYALVLOQKVTDATQIVVCLLSNRKDYD 483
DB 510 SLIQNLFKVTPAMGQMRKAIMLEVDDRTYALVLOQKVTDATQIVVCLLSNRKDYD 569
QY 484 AIKKYLCTDPTSCQCVARTLGKQOTVMAIATKIALQMNCKMGELWVDIPLKLVIV 543
DB 570 AIKKYLCTDPTSCQCVARTLGKQOTVMAIATKIALQMNCKMGELWVDIPLKLVIV 629
QY 544 GIDCYHDMTAGRSIAGFVASINEGTRWFSRCIFQDRGOELVDGLKVCLOAALRAWNSC 603
DB 630 GIDCYHDMTAGRSIAGFVASINEGTRWFSRCIFQDRGOELVDGLKVCLOAALRAWNSC 689
QY 604 NEYMPRIIYVRDVGQGLKTLVNYEVPQFLDCLKISGRGNPRLTVIVVKRVNTRFF 663
DB 690 NEYMPRIIYVRDVGQGLKTLVNYEVPQFLDCLKISGRGNPRLTVIVVKRVNTRFF 749
QY 664 AQSGRLQNPFGTVIDVEVTRPEWYDFIVSQAVRSVSPHYNVIYDNSGLKPDHIQ 723
DB 750 AQSGRLQNPFGTVIDVEVTRPEWYDFIVSQAVRSVSPHYNVIYDNSGLKPDHIQ 809
QY 724 RLTYKLCHIIYNNWPGVIRVPAPQYAHKLAFLVGQSIHREPNSLSNRLYYL 775
DB 810 RLTYKLCHIIYNNWPGVIRVPAPQYAHKLAFLVGQSIHREPNSLSNRLYYL 861

RESULT 3
Q8TBYS ID Q8TBYS PRELIMINARY; PRT; 861 AA.
AC Q8TBYS;
DT 01-JUN-2002 (TrEMBLrel. 21, Created)
DT 01-JUN-2002 (TrEMBLrel. 21, Last sequence update)
DE Piwi-like 1 (Drosophila).
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]_
RP SEQUENCE FROM N.A.
```

RC TISSUE-Testis;
RA Strausberg R.;
RL Submitted (APR-2002) to the EMBL/GenBank/DBJ databases.
DR EMBL; BC028581; AA28581.1; -
DR InterPro; IPR003100; PAZ.
DR InterPro; IPR003165; Piwi.
DR Pfam; PF02170; PAZ; 1.
DR Pfam; PF02171; Piwi; 1.
DR PROSITE; PSS0821; PAZ; 1.
DR PROSITE; PSS0822; PIWI; 1.
SQ SEQUENCE 861 AA; 98545 MW; D33376BEDEDF743A CRC64;

Query Match 99.3%; Score 4072; DB 4; Length 861;
Best Local Similarity 99.6%; Pred. No. 8.4e-316;
Matches 769; Conservative 1; Mismatches 2; Indels 0; Gaps 0;
QY 4 GVNTRQNLHVKESTGSSGIIVRLSTNNHRLTSRQWALYQYHIDYNPLMEARRLSAL 63
DB 90 GVNTRQNLHVKESTGSSGIIVRLSTNNHRLTSRQWALYQYHIDYNPLMEARRLSAL 149
QY 64 LFOHEDLIGKCHAFDGTILFLPKRLQKQVTEVFSKTRNGEDVRITITLTNELPPTSPCTL 123
DB 150 LFOHEDLIGKCHAFDGTILFLPKRLQKQVTEVFSKTRNGEDVRITITLTNELPPTSPCTL 209
QY 124 QFYNIIFRLLKIMNLQOIGRNYNPNPIDIPSHRLVWPFGTTSILQYENSIMLCTDV 183
DB 210 QFYNIIFRLLKIMNLQOIGRNYNPNPIDIPSHRLVWPFGTTSILQYENSIMLCTDV 269
QY 184 SHKVLRSVLDPMFNFYHQTTEHKFQOVSKELIGLVLTKNKTYRVDIDWDQNP 243
DB 270 SHKVLRSVLDPMFNFYHQTTEHKFQOVSKELIGLVLTKNKTYRVDIDWDQNP 329
QY 244 STFKKADGSEVSFLEYRKYQNOEITDLKQVLSQPKRRRPGGTLPGPAMLIPELCVL 303
DB 330 STFKKADGSEVSFLEYRKYQNOEITDLKQVLSQPKRRRPGGTLPGPAMLIPELCVL 389
QY 304 TGLTDKMRDNFNKDLAVHTRLTPEQRQREVGRLLIDYIHKNQNVQRELWDGLSFDNL 363
DB 390 TGLTDKMRDNFNKDLAVHTRLTPEQRQREVGRLLIDYIHKNQNVQRELWDGLSFDNL 449
QY 364 LSFSGRILQTEKHOGGKTFDYNQPADWSKETRGAPLISVKPLDNWLLIYTRRNYEAA 423
DB 450 LSFSGRILQTEKHOGGKTFDYNQPADWSKETRGAPLISVKPLDNWLLIYTRRNYEAA 509
QY 424 SLIQNLFKVTPANGQMKKAIMIEVDRTTEAYLRVLQKQVTDQTVVCLLSNRKDKYD 483
DB 510 SLIQNLFKVTPANGQMKKAIMIEVDRTTEAYLRVLQKQVTDQTVVCLLSNRKDKYD 569
QY 484 AIKKYLCTDPTSPQCVVARTLGKQOTVMAITKIALQVNCXGGLMRVDIPLKLVMI 543
DB 570 AIKKYLCTDPTSPQCVVARTLGKQOTVMAITKIALQVNCXGGLMRVDIPLKLVMI 629
QY 544 GIDCVHDMTAGRSIAGFVASINEGTRMFSFCIFQDRGOELVDGLKVCLOALRAWNSC 603
DB 630 GIDCVHDMTAGRSIAGFVASINEGTRMFSFCIFQDRGOELVDGLKVCLOALRAWNSC 689
QY 604 NEYMPRIIVYRDGVGDGLKTLVNYEVPQFLDCLKSIQGRYNPRLTIVLVKKRVNTRFF 663
DB 690 NEYMPRIIVYRDGVGDGLKTLVNYEVPQFLDCLKSIQGRYNPRLTIVLVKKRVNTRFF 749
QY 664 AQSGRLQNLPGTVIDEVTPEWDFPIVQAVRSQSVSPTHYNYVDSGLKPDHIQ 723
DB 750 AQSGRLQNLPGTVIDEVTPEWDFPIVQAVRSQSVSPTHYNYVDSGLKPDHIQ 809
QY 724 RLTYKLCHYYNPGVIRVPACQVAHKLAFVVGSIHREPNSLSNRLIYL 775
DB 810 RLTYKLCHYYNPGVIRVPACQVAHKLAFVVGSIHREPNSLSNRLIYL 861

RESULT 4
ID O95404 PRELIMINARY; PRT; 861 AA.
AC O95404;

DT 01-MAY-1999 (TrEMBLrel. 10, Created)
DT 01-MAR-2002 (TrEMBLrel. 20, Last sequence update)
DT 01-OCT-2002 (TrEMBLrel. 22, Last annotation update)
DE HIWI.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]_TaxID=9606;
RP SEQUENCE FROM N.A.
RC TISSUE=Testis;
RX MEDLINE=99069219; PubMed=9851978;
RA Cox D.N., Chao A., Baker J., Chang L., Qiao D., Lin H.;
RT "A novel class of evolutionarily conserved genes defined by piwi are
RL essential for stem cell self-renewal."; Genes Dev. 12:3715-3727 (1998).
RN [2]
RP SEQUENCE FROM N.A.
RC TISSUE=Testis;
RA Qiao D., Zeeman A.-M., Deng W., Looijenga L.H.J., Lin H.;
RT "Molecular characterization of hiwi, a human member of the piwi stem
RL cell gene family whose overexpression is correlated to seminomas."; Submitted (JAN-2002) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF104260; AAC97371.2; -
DR Genew; HGNC:9007; PIWIL1.
DR InterPro; IPR003100; PAZ.
DR InterPro; IPR003165; Piwi.
DR Pfam; PF02170; PAZ; 1.
DR Pfam; PF02171; Piwi; 1.
DR PROSITE; PSS0821; PAZ; 1.
DR PROSITE; PSS0822; PIWI; 1.
SQ SEQUENCE 861 AA; 98530 MW; 43D7F60E99D997B7 CRC64;
Query Match 99.2%; Score 4071; DB 4; Length 861;
Best Local Similarity 99.7%; Pred. No. 1e-315;
Matches 770; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
QY 4 GVNTRQNLHVKESTGSSGIIVRLSTNNHRLTSRQWALYQYHIDYNPLMEARRLSAL 63
DB 90 GVNTRQNLHVKESTGSSGIIVRLSTNNHRLTSRQWALYQYHIDYNPLMEARRLSAL 149
QY 64 LFOHEDLIGKCHAFDGTILFLPKRLQKQVTEVFSKTRNGEDVRITITLTNELPPTSPCTL 123
DB 150 LFOHEDLIGKCHAFDGTILFLPKRLQKQVTEVFSKTRNGEDVRITITLTNELPPTSPCTL 209
QY 124 QFYNIIFRLLKIMNLQOIGRNYNPNPIDIPSHRLVWPFGTTSILQYENSIMLCTDV 183
DB 210 QFYNIIFRLLKIMNLQOIGRNYNPNPIDIPSHRLVWPFGTTSILQYENSIMLCTDV 269
QY 184 SHKVLRSVLDPMFNFYHQTTEHKFQOVSKELIGLVLTKNKTYRVDIDWDQNP 243
DB 270 SHKVLRSVLDPMFNFYHQTTEHKFQOVSKELIGLVLTKNKTYRVDIDWDQNP 329
QY 244 STFKKADGSEVSFLEYRKYQNOEITDLKQVLSQPKRRRPGGTLPGPAMLIPELCVL 303
DB 330 STFKKADGSEVSFLEYRKYQNOEITDLKQVLSQPKRRRPGGTLPGPAMLIPELCVL 389
QY 304 TGLTDKMRDNFNKDLAVHTRLTPEQRQREVGRLLIDYIHKNQNVQRELWDGLSFDNL 363
DB 390 TGLTDKMRDNFNKDLAVHTRLTPEQRQREVGRLLIDYIHKNQNVQRELWDGLSFDNL 449
QY 364 LSFSGRILQTEKHOGGKTFDYNQPADWSKETRGAPLISVKPLDNWLLIYTRRNYEAA 423
DB 450 LSFSGRILQTEKHOGGKTFDYNQPADWSKETRGAPLISVKPLDNWLLIYTRRNYEAA 509
QY 424 SLIQNLFKVTPANGQMKKAIMIEVDRTTEAYLRVLQKQVTDQTVVCLLSNRKDKYD 483
DB 510 SLIQNLFKVTPANGQMKKAIMIEVDRTTEAYLRVLQKQVTDQTVVCLLSNRKDKYD 569
QY 484 AIKKYLCTDPTSPQCVVARTLGKQOTVMAITKIALQVNCXGGLMRVDIPLKLVMI 543
DB 570 AIKKYLCTDPTSPQCVVARTLGKQOTVMAITKIALQVNCXGGLMRVDIPLKLVMI 629

```

QY 544 GIDCYHDMTAGRSIAGFVASINEGTRWFSRCIFODRGQELVDGLKVCLOALRAWNSC 603
DB 630 GIDCYHDMTAGRSIAGFVASINEGTRWFSRCIFODRGQELVDGLKVCLOALRAWNSC 689
QY 604 NEYMPRIIVYRDGVGDGQKLTIVNVEVPOFLDCLKSIGRGYNPRLTVIVVKRVNTRFF 663
DB 690 NEYMPRIIVYRDGVGDGQKLTIVNVEVPOFLDCLKSIGRGYNPRLTVIVVKRVNTRFF 749
QY 664 AOSGGLQNPFGTVIDVEVTRPEWYDFIVSQVRSVSPHYNVIYDNSGLKPDHIQ 723
DB 750 AOSGGLQNPFGTVIDVEVTRPEWYDFIVSQVRSVSPHYNVIYDNSGLKPDHIQ 809
QY 724 RLTYKLCHIVYNNWPGVIRVPAPCOYAHKLAFLVQSIHREPNSLSNRLYYL 775
DB 810 RLTYKLCHIVYNNWPGVIRVPAPCOYAHKLAFLVQSIHREPNSLSNRLYYL 861

RESULT 5
Q9JMB7
ID Q9JMB7 PRELIMINARY; PRT; 862 AA.
AC Q9JMB7
DT 01-OCT-2000 (TrEMBLrel. 15, Created)
DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DE 01-OCT-2002 (TrEMBLrel. 22, Last annotation update)
DE MIWI (piwi).
GN PIWILL OR MIWI.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RA Miyagawa S.K., Kimura T., Nakano T.;
RT "Molecular Cloning and characterization of Piwi family genes.";
RL Submitted (SEP-1999) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RC TISSUE=Testis;
RA Deng W., Lin H.;
RT "Miwi, a murine homolog of piwi, encodes a cytoplasmic protein
RT essential for spermatogenesis.";
RL Submitted (OCT-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; AB032604; BAA93705.1; -
DR EMBL; AF438405; AAL31014.1; -
DR MGD; MGI:1928897; Piwi1.
DR InterPro; IPR003100; PAZ.
DR InterPro; IPR003165; Piwi.
DR Pfam; PF02170; PAZ; 1.
DR Pfam; PF02171; Piwi; 1.
DR PROSITE; PS50821; PAZ; 1.
DR PROSITE; PS50822; PIWI; 1.
SQ SEQUENCE 862 AA; 98574 MW; 45588D13284CCC4C CRC64;

Query Match 97.0%; Score 3980; DB 11; Length 862;
Best Local Similarity 96.4%; Pred. No. 1.8e-308;
Matches 744; Conservative 19; Mismatches 9; Indels 0; Gaps 0;

QY 4 GYNTQNLDHVKESTGSSGIIVRLTSRPNALYQYHIDYPLMEARLRSAL 63
DB 91 GYNTQNLDHVKESTGSSGIIVRLTSRPNALYQYHIDYPLMEARLRSAL 150
QY 64 LFOHEDLIGKCHAFDGTILFLPKRLOQKVTVEFSQTRNGEDVRIITLTNELPPTSPCL 123
DB 151 LFOHEDLIGKCHAFDGTILFLPKRLOQKVTVEFSQTRNGEDVRIITLTNELPPTSPCL 210
QY 124 QFYNIIFRLLKIMNLQIQGRYNNPNPDIDPSHRLVWPQFTTSIIQYENSIMCLTDV 183
DB 211 QFYNIIFRLLKIMNLQIQGRYNNPNPDIDPSHRLVWPQFTTSIIQYENSIMCLTDV 270
QY 184 SHKVLRSVLDPMFNFYHQTEHKFQGVSKELIGLVLTLYNNKXRVDDIDWDQNP 243
DB 271 SHKVLRSVLDPMFNFYHQTEHKFQGVSKELIGLVLTLYNNKXRVDDIDWDQNP 330

```

```

QY 244 STFKKADSGSEVSFLYIRKQYNQEIITDLKQPVLSQPKRRRGGTLPGFAMLIPELCYL 303
DB 331 STFKKADSGSEVSFLYIRKQYNQEIITDLKQPVLSQPKRRRGGTLPGFAMLIPELCYL 390
QY 304 TGLTDKMRDNFNWMDLAVHTRLTPEQREVRGLLDIYHKNDNVQRELRDGLSFDNSL 363
DB 391 TGLTDKMRDNFNWMDLAVHTRLTPEQREVRGLLDIYHKNDNVQRELRDGLSFDNSL 450
QY 364 LSFSGRIILQTEKHOGGKTFDYNPOFADWSKETRCAPLISVAPLDNWLIIYTRRYEAA 423
DB 451 LSFSGRIILQSEKHOGGKTFDYNPOFADWSKETRCAPLISVAPLDNWLIIYTRRYEAA 510
QY 424 SLIQNLKFKVTPAMGNQMKKAIMIEVDDRETEAYLRVLOQKVTDATQIVVCLSSNRKQYD 483
DB 511 SLIQNLKFKVTPAMGIQMKKAIMIEVDDRETEAYLRVLOQKVTSDTQIVVCLSSNRKQYD 570
QY 484 AIKKYLCVDCPTSPSOVARTILGKQOTVMAATKIALQNCXMGGLWRVDIPLKLMIV 543
DB 571 AIKKYLCVDCPTSPSOVARTILGKQOTVMAATKIALQNCXMGGLWRVDIPLKLMIV 630
QY 544 GIDCYHDMTAGRSIAGFVASINEGTRWFSRCIFODRGQELVDGLKVCLOALRAWNSC 603
DB 631 GIDCYHDMTAGRSIAGFVASINEGTRWFSRCIFODRGQELVDGLKVCLOALRAWNSC 690
QY 604 NEYMPRIIVYRDGVGDGQKLTIVNVEVPOFLDCLKSIGRGYNPRLTVIVVKRVNTRFF 663
DB 691 NEYMPRIIVYRDGVGDGQKLTIVNVEVPOFLDCLKSIGRGYNPRLTVIVVKRVNTRFF 750
QY 664 AOSGGLQNPFGTVIDVEVTRPEWYDFIVSQVRSVSPHYNVIYDNSGLKPDHIQ 723
DB 751 AOSGGLQNPFGTVIDVEVTRPEWYDFIVSQVRSVSPHYNVIYDNSGLKPDHIQ 810
QY 724 RLTYKLCHIVYNNWPGVIRVPAPCOYAHKLAFLVQSIHREPNSLSNRLYYL 775
DB 811 RLTYKLCHIVYNNWPGVIRVPAPCOYAHKLAFLVQSIHREPNSLSNRLYYL 862

```

RESULT 6

```

Q8NA60
ID Q8NA60 PRELIMINARY; PRT; 829 AA.
AC Q8NA60;
DT 01-OCT-2002 (TrEMBLrel. 22, Created)
DT 01-OCT-2002 (TrEMBLrel. 22, Last sequence update)
DT 01-WAR-2003 (TrEMBLrel. 23, Last annotation update)
DE Hypothetical protein FLJ35814 (piwi).
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Testis;
RA Oshima A., Takahashi-Fujii A., Tanase T., Imose N., Takeuchi K.,
RA Arita M., Musashino K., Yuuki H., Hara H., Sugiyama T., Irie R.,
RA Otsuki T., Sato H., Wakamatsu A., Ishii S., Yamamoto J., Isono Y.,
RA Kawai-Hio Y., Saito K., Nishikawa T., Kimura K., Yamashita H.,
RA Matsuo K., Nakamura Y., Sekine K., Kikuchi H., Kanda K., Wagatsuma M.,
RA Murakawa K., Kanehori K., Sugiyama A., Kawakami B., Suzuki Y.,
RA Sugano S., Nagahori K., Masuho Y., Nagai K., Isogai T.;
RT "NEDO human cDNA sequencing project.";
RL Submitted (JUL-2002) to the EMBL/GenBank/DBJ databases.
DR EMBL; AK093133; BAC04068.1; -
DR InterPro; IPR003100; PAZ.
DR InterPro; IPR003165; Piwi.
DR Pfam; PF02170; PAZ; 1.
DR Pfam; PF02171; Piwi; 1.
DR PROSITE; PS50821; PAZ; 1.
DR PROSITE; PS50822; PIWI; 1.
KW Hypothetical protein.
SQ SEQUENCE 829 AA; 94816 MW; 3AF9FD1D0FA9F5A7 CRC64;

```

Query Match 94.7%; Score 3883; DB 4; Length 829;
 Best Local Similarity 99.9%; Pred. No. 9.5e-301;

01-MAR-2003 (Tremblrel..23, Last annotation update)
 Similar to piwi like homolog 1 (Drosophila).

02-Homo sapiens (Human)
 03-Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 04-Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 05-NCBI_TaxID=9606;
 06-[1]_TaxID=9606;
 07-SEQUENCE FROM N.A.
 08-SEQUENCE=Testis;
 09-Strausberg R.;
 10-Submitted (JUN-2002) to the EMBL/GenBank/DBJ databases.
 11-EMBL; BC031060.1; -
 12-InterPro; IPR003100; PAZ.
 13-InterPro; IPR003165; Piwi.
 14-Pfam; PF02170; PAZ; 1.
 15-Pfam; PF02171; Piwi; 1.
 16-PROSITE; PS0821; PAZ; 1.
 17-PROSITE; PS0822; PIWI; 1.
 18-SEQUENCE 852 AA; 96588 MW; 37769E078B96D13 CRC64;

Query Match 51.3%; Score 2103.5; DB 4; Length 852;
 Best Local Similarity 51.2%; Pred. No. 9.3e-159;
 Matches 398; Conservative 146; Mismatches 225; Indels 9; Gaps 4;

QY 1 MIFGVTNQLNDHVKESKTSSGIIVRLSTNHFRLTSPQWALYQYHIDYNPLMEARLR 60
 DB 81 MDLSICTREKLAHVNRCKTGSSGIPVKLVNLFNLPDQWQLYQYHVTIPDLASRLR 140
 QY 61 SALLFQHEDLIGKCHAFDGTILFLPKRLOQKVEVFSKTRNGEDVRITITLNLPTSP 120
 DB 141 IALLYSHSELNSKAKAFDGAIFLSQLEKVEVFSKTRNGEDVRITITLNLPTSP 200
 QY 121 TGLQYNIIFRLLKIMNLQOIGRNYNPNDDIPSHRLVWPGFTTSILQYENSIMLC 180
 DB 201 VCIOVFNIIFRLLKIMNLQOIGRNYNPNDDIPSHRLVWPGFTTSILQYENSIMLC 260
 QY 181 TDVSHKVLRSVLDPMFNFIHTEBEKFOVSKELIGLVLTKNKTYRVDIDWDQ 240
 DB 261 ADVSKVLRSVLDPMFNFIHTEBEKFOVSKELIGLVLTKNKTYRVDIDWDQ 320
 QY 241 NPKSTFKKADGSEVSEFLYRKYQYQYQYQYQYQYQYQYQYQYQYQYQYQYQYQY 360
 DB 321 KPTHTFQKRDGTETIYVYQYQYQYQYQYQYQYQYQYQYQYQYQYQYQYQY 439
 QY 361 SNLLSFGRILOTEKIHOGGKTFDYNPQ---FADWSKETRGAPLISVKPLDNWLLIYTR 417
 DB 440 SQ-ISLTGRIVPSEKILMQ---DHICQPVSAADWSKDRTCKILNAQSLNTWLLCSDR 494
 QY 418 NYEAANSILQNLFPKVTAPMGQMKRAIMIEVDRTETAYLRVLOQKVTADTQIVVCLASN 477
 DB 495 TEYVAESPLNCLRRVAGSGMGNVDPYKIKYQENPAAFPVRAIQYVDVQVAVCILPSN 554
 QY 478 RKDYADALKYLCCTCPSPSCVARTLGKQQTWATKIALQNKQMGELNRVIDPL 537
 DB 555 QRTYDSIKKYSDDCPVSPSCVARTLGKQQTWATKIALQNKQMGELNRVIDPL 614
 QY 538 KLVMIYDIDYHDMTAGRRSTAGFVASINEGRTWFSRCIFQDRGOELVDGLKVCLOAAL 597
 DB 615 KSLAVVGIDVCKDALSKDVVGVGVAVNPRITRFSICILQRTWVDADCLKVFMTGAL 674
 QY 598 RAWNSCNEYMPERIIYVYRGGDGLKTLVNYVPPQFDDCLKSIGRGNPLRTIVVVKR 657
 DB 675 NKWYKYNHDLPIIYVYRAGVGGDGLKTLVNYVPPQFDDCLKSIGRGNPLRTIVVVKR 734
 QY 658 VNTFFAQSGRLQNLPGTVDIVETWPEYDFIVSQVRSVSGPHTYNYIDNSGL 717
 DB 735 CMPREFFEMNRVQNPPLGTVDIVETWPEYDFIVSQVRSVSGPHTYNYIDNSGL 794
 QY 718 KPDHQRITLYKLCYIYNNWPGVIRVPAPQYAHKLAFLVGSQIHRPNLSNRLYL 775

DB 735 KPDHQRITLYKLCYIYNNWPGVIRVPAPQYAHKLAFLVGSQIHRPNLSNRLYL 852

RESULT 9
 Q8N9V8

ID Q8N9V8 PRELIMINARY; PRT; 852 AA.
 AC Q8N9V8;
 DT 01-OCT-2002 (Tremblrel..22, Created)
 DT 01-OCT-2002 (Tremblrel..22, Last sequence update)
 DT 01-MAR-2003 (Tremblrel..23, Last annotation update)
 DE Hypothetical protein FLJ36156 (piwi).
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 OX NCBI_TaxID=9606;
 RN [1]_TaxID=9606;
 RP SEQUENCE FROM N.A.
 RC TISSUE=Testis;
 RA Ishibashi T., Kanehori K., Yosida M., Watanabe S., Ishida S., Ono Y.,
 RA Hotuta T., Hiraoka S., Murakawa K., Takiguchi S., Kusano J., Chiba Y.,
 RA Watanabe M., Fujimori K., Tanai H., Ishida M., Yamashita H., Chiba Y.,
 RA Sugiyama T., Irie R., Otsuki T., Sato H., Wakamatsu A., Ishii S.,
 RA Yamamoto J., Isono Y., Kawai-Hio Y., Saito K., Nishikawa T.,
 RA Kimura K., Matsuo K., Nakamura Y., Sekine M., Kikuchi H., Kanda K.,
 RA Wagatsuma M., Takahashi-Fujii A., Oshima A., Sugiyama A., Kawakami B.,
 RA Suzuki Y., Sugano S., Nagahari K., Masuho Y., Nagai K., Isogai T.;
 RT "NEDO human cDNA sequencing project";
 RL Submitted (JUL-2002) to the EMBL/GenBank/DBJ databases.
 DR EMBL; AK093475; BAC04179.1; -
 DR InterPro; IPR003100; PAZ.
 DR InterPro; IPR003165; Piwi.
 DR Pfam; PF02170; PAZ; 1.
 DR Pfam; PF02171; Piwi; 1.
 DR PROSITE; PS0821; PAZ; 1.
 DR PROSITE; PS0822; PIWI; 1.
 KW Hypothetical protein.
 SQ SEQUENCE 852 AA; 96560 MW; 75C6EFA4E70701B CRC64;

Query Match 51.2%; Score 2101.5; DB 4; Length 852;
 Best Local Similarity 51.0%; Pred. No. 1.3e-158;
 Matches 397; Conservative 148; Mismatches 224; Indels 9; Gaps 4;

QY 1 MIFGVTNQLNDHVKESKTSSGIIVRLSTNHFRLTSPQWALYQYHIDYNPLMEARLR 60
 DB 81 MDLSICTREKLAHVNRCKTGSSGIPVKLVNLFNLPDQWQLYQYHVTIPDLASRLR 140
 QY 61 SALLFQHEDLIGKCHAFDGTILFLPKRLOQKVEVFSKTRNGEDVRITITLNLPTSP 120
 DB 141 IALLYSHSELNSKAKAFDGAIFLSQLEKVEVFSKTRNGEDVRITITLNLPTSP 200
 QY 121 TGLQYNIIFRLLKIMNLQOIGRNYNPNDDIPSHRLVWPGFTTSILQYENSIMLC 180
 DB 201 VCIOVFNIIFRLLKIMNLQOIGRNYNPNDDIPSHRLVWPGFTTSILQYENSIMLC 260
 QY 181 TDVSHKVLRSVLDPMFNFIHTEBEKFOVSKELIGLVLTKNKTYRVDIDWDQ 240
 DB 261 ADVSKVLRSVLDPMFNFIHTEBEKFOVSKELIGLVLTKNKTYRVDIDWDQ 320
 QY 241 NPKSTFKKADGSEVSEFLYRKYQYQYQYQYQYQYQYQYQYQYQYQYQYQYQYQY 360
 DB 321 KPTHTFQKRDGTETIYVYQYQYQYQYQYQYQYQYQYQYQYQYQYQYQYQY 379
 QY 301 CYLTGLTDQATSDPQMKAVAEKTRLSPSGRQRLARLVNIOQNTNAREFEETWGLHFG 439
 DB 380 CFLTGLTDQATSDPQMKAVAEKTRLSPSGRQRLARLVNIOQNTNAREFEETWGLHFG 494
 QY 361 SNLLSFGRILOTEKIHOGGKTFDYNPQ---FADWSKETRGAPLISVKPLDNWLLIYTR 417
 DB 440 SQ-ISLTGRIVPSEKILMQ---DHICQPVSAADWSKDRTCKILNAQSLNTWLLCSDR 494
 QY 418 NYEAANSILQNLFPKVTAPMGQMKRAIMIEVDRTETAYLRVLOQKVTADTQIVVCLASN 477

Db 495 TEVVAESFLNCLRRVAGSMGFNVDPKIKVOENPAFVRAIQYVDPDQVLMCILPSN 554
 QY 478 RDKYDAIKKYLCTDCTPSPQCVARTLGGKQVMAIAIKALQNCCKWGGLWRVDEPL 537
 Db 555 QTYDYSDIKYLISSDCPVSPQCVARTLGGKQVMAIAIKALQNCCKWGGLWRVDEPL 614
 QY 538 KLVMIVGIDCYHDMTAGRSIAGFVASINEGTRWFSRCIFQDRGQELVDGLKVLQAL 597
 Db 615 KSLVVGIDVCKDALSXDVMVGVCAVNPRTIRWFSRCILQRTMTDADCLKVFMTGAL 674
 QY 598 RAWNSCNEWPSRIIVYRGVGDGOLKTLVNVYVPOFLDCLKSLRGYNPBLTVLVKKR 657
 Db 675 NKWYKINDLPARIIVYRGVGDGOLKTLVNVYVPOFLDCLKSLRGYNPBLTVLVKKR 734
 QY 658 VNTTFAQSGRGLQPLGCTVIDVEVTRPFWYDFVVSQVRSVSPHYNYVDSNGL 717
 Db 735 CNPRFFTEMRTVQNPPLGTVDSSEATRNWDYFLISQVACRGTVSPHYNYVDSNGL 794
 QY 718 KPDHQRILTYKLCHYYNWPVIRVPAPCOYAHKLAFLVQSGIHRPNSLRNLYYL 775
 Db 795 KPDHQRILTYKLCHYYNWPVIRVPAPCOYAHKLAFLVQSGIHRPNSLRNLYYL 852

RESULT 10

Q9GPA8 PRELIMINARY; PRT; 854 AA.
 AC Q9GPA8;
 DT 01-MAR-2001 (TREMELrel. 16, Created)
 DT 01-MAR-2001 (TREMELrel. 16, Last sequence update)
 DT 01-OCT-2002 (TREMELrel. 22, Last annotation update)
 DE Seawi.
 OS Strongylocentrotus purpuratus (Purple sea urchin).
 OC Eukaryota; Metazoa; Echinodermata; Eleutherozoa; Echinozoa;
 OC Echinoidea; Euechinoidea; Echinacea; Echinoidea; Strongylocentrotidae;
 OC Strongylocentrotus.
 NCBI_TaxID=7668;
 RN [1]
 RP SEQUENCE FROM N.A.
 RA Rodriguez A.J., Sonder E.M.;
 RT "Seawi - Cloning and Molecular Characterization of a Sea Urchin
 RT Homolog of Pw1."
 RL Submitted (NOV-2000) to the EMBL/GenBank/DBJ databases.
 DR EMBL; AY014899; AAG42533.1; --
 DR InterPro; IPR003100; PAZ.
 DR InterPro; IPR003165; Piwi.
 DR Pfam; PF02170; PAZ; 1.
 DR Pfam; PF02171; Piwi; 1.
 DR PROSITE; P50821; PAZ; 1.
 DR PROSITE; P50822; PIWI; 1.
 SQ SEQUENCE 854 AA; 96722 MW; 509A1D39C0D1922C CRC64;

Query Match 49.0%; Score 2009.5; DB 5; Length 854;
 Best Local Similarity 50.5%; Pred. No. 2.9e-151;
 Matches 386; Conservative 133; Mismatches 236; Indels 9; Gaps 6;
 QY 14 VKESKTGSGGIIVRLSTNHFRLTSRQWALYQVHDYDYNPLMEARRLSALLFQHEDLICK 73
 Db 98 VKQAIAGDK---IALTANGFKLTKTDQWLYQVYRVDFEPEILNPRAPFALLAGHSGALLGK 154
 QY 74 CHAFDGTILFLPKRLOQKTEVFSKTRNGEDVRITITLNEIPPTSPCTCLQFYNIIFRL 133
 Db 155 GLTLDMDTVSLFKLAEKVTNLGAERKGSNNVCVTHVATLNPAPNPLHLNVLFRRC 214
 QY 134 LKIMLQIQIRNNYVNPNDIPSHRLVTPGFTTTSILOVENSIMLCTDVSHKVLRSVTV 193
 Db 215 LKMINMEQVGRNYDPTAIDIKQHLQWLPFVTSILOVEDVWLLSDISHKVLRTQV 274
 QY 194 LDFNFNYHOTEHFKQEQSVKSELIGLVLTLYNNKTYRVDDIDWDPNPKSTFKKADGSE 253
 Db 275 WEVANDLFNKA-RGRPKETIKLMIQVILTKYNNKTYRVDDIDFDTTADTFETRSG-P 332
 QY 254 VSLEYRYKQYNEITLQKPLVLSQPKRRRGGGTLPQGMILIPLCVLTGLTDMKND 313

Db 333 VSYVDYFKSYERYIHDVNPQMLVSRPKREKKGV--GPAYLPPELCELTLGLSDMRAD 390
 QY 314 ENVMKDLAVHTRITPBOREVRGLIDYTHKNDVQRELRDWSLSDNLLSFSRILQT 373
 Db 391 FNVKDMGQHTRVGPDRCRTLSGFIKLSNEEVKTYLDSWGEFDEKQVLTGVLFP 450
 QY 374 EKHQGGKTFDYNPOPADWSKETRGAPLISVLPDLNLLIYTRRYEANSIIONLFKVT 433
 Db 451 EKLFGSKGQFSYNPSNADWSRDRGNALTDKILANNKIFYTRRDANRGQDFIKSLRVA 510
 QY 434 PANGMOMKAIMLEV--DDETEAVLRVLOQKVTADTQIVVCLLSSNRKDYDAIKKYLCTD 492
 Db 511 NPMGMNVGRPEIVEIPDDRTETITTSLOQIAQDTQIVVILPTNRKDYDAIKKTCVTV 570
 QY 493 CPTPSCVIVARTLGGKQVMAIAIKALQNCCKWGGLWRVDEPLKLVNIVGIDCYHDMT 552
 Db 571 HPCPSQVIVSRTLSKQMLMSVATKIAMQNCCKWGGLWRVDEPLSNLMIIGDSDHSL 630
 QY 553 AGERSIAGFVASINEGTRWFSRCIFQDRGQELVDGLKVLQALQNCCKWGGLWRVDEPL 612
 Db 631 TKGSRVLGFGVSNMKSQTSFSSCAPQHAQGFGANLSTLMNNALKRYTYQINEKPFERII 690
 QY 613 VYRDGVGDGOLKTLVNVYVPOFLDCL-KSIGRGYNPRLTVIVVKKRVNTRFFAQSGRIQ 671
 Db 691 IFEDGVGDSQVNLVDVYELKQIKDYLKVYPOGTVHKLAIVVVKKINRFFANLGGLS 750
 QY 672 NPLPGTVIVDEVTRPFWYDFVVSQVRSVSPHYNYVDSNGLKPDHIQRLTYKLCH 731
 Db 751 NPPEGTVIDDVVTKPHLYDYFIISQSVRQSGVSPSYNVVYDTTGLKPDHQRILTYKLTH 810
 QY 732 IYVNPQVIRVPAPCOYAHKLAFLVQSGIHRPNSLRNLYYL 775
 Db 811 LYFNWPGTVRVPAPCOYAHKLAFLVQSGIHRPNSLRNLYYL 854

RESULT 11
 Q8CGT6 PRELIMINARY; PRT; 878 AA.
 AC Q8CGT6;
 DT 01-MAR-2003 (TREMELrel. 23, Created)
 DT 01-MAR-2003 (TREMELrel. 23, Last sequence update)
 DT 01-MAR-2003 (TREMELrel. 23, Last annotation update)
 DE Miwi 2 protein.
 GN Miwi2.
 OS Mus musculus (Mouse).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 NCBI_TaxID=10090;
 RN [1]
 RP SEQUENCE FROM N.A.
 RA Carmell M.A., Xuan Z., Zhang M.Q., Hannon G.J.;
 RT "The Argonaute family: tentacles that reach into RNAi, developmental
 RT control, stem cell maintenance, and tumorigenesis."
 RL Genes Dev. 16:2733-2742(2002).
 RN [2]
 RP SEQUENCE FROM N.A.
 RA Adams M., Mural R.;
 RL Submitted (JUL-2002) to the EMBL/GenBank/DBJ databases.
 DR EMBL; AY135692; AAN75583.1; -- 3C6265E0E4ED7C3F CRC64;
 SQ SEQUENCE 878 AA; 98617 MW; 3C6265E0E4ED7C3F CRC64;

Query Match 45.3%; Score 1859.5; DB 11; Length 878;
 Best Local Similarity 46.5%; Pred. No. 2.8e-139;
 Matches 370; Conservative 134; Mismatches 209; Indels 83; Gaps 10;
 QY 4 GVNTRQMLDHWKSGKSGSGIIVRLSTNHFRLTSRQWALYQVHDYDYNPLMEARRLSAL 63
 Db 142 GVCRTREKLTWKDCKTGSSGIPVRLVTLNPLNLDLPQDWLYQVHYTVSPDLASRRRLAL 201
 QY 64 LFOHEDLIGKCHAFDGTILFLPKRLOQKTEVFSKTRNGEDVRITITLNEIPPTSPCTCL 123
 Db 202 LYNHSLSDKAKAPDAGSLFLSEKLDQKVTELTSTQRTETIKITLITLSKLPFNPVCI 261

```

QY 124 QFVNIIFRELLKIMNLQOIGRNYNPNPDIPIPSHRLVWPFGTTSLIQYENSIMLCITDV 183
DB 262 QFNVIFRILKIMLSYQIGRNYFSEVEIP-----QY-NKLFNADV 305
QY 184 SHKVLSETVLDPMFNFYHQTTEHKFQEQVSKELIGLVLTIKYNNKTYRVDIDWDQNPX 243
DB 306 NYKVLNSETVLDPMFTDCLRTGNSCFTEMCHQOLGVLVLTIRYNNKTYRIDDDWSVKPT 365
QY 244 STFKKADGSEVSEFLEYYRYKYNQBEITDLKQPVLVGQPKRRRPGGTLPGPAMLPDEL CYL 303
DB 366 QAFQKDGSEVTVYDYKQYDITLSDNQPVLSLKKRNDNSE-FQWHLMPDELCEL 424
QY 304 TGLTDMRNDENOMKDLAVHRLTPQORQREYGRILDIYHKNDNVQRELRD-----N-- 355
DB 425 TGLSSQATSDFLMKAVAEETRLSPVGRQOQLARLVDDIORTLPSSQEVLSHTSLPLWAP 484
QY 356 ---GLSFD---SMLSFSGRILOTEKHOG-----GKTFDYNPQF-ADWSKETRGA 399
DB 485 EPGGLSAPLSTVLPFAOQLLTALSFGIPLPHLKPPSFLFLCQAPAFADWSKDMRSC 544
QY 400 PLISVKEPLDNWLLIYTRNVEAANSILQNLKVTYPAMGMQMKAMIEVDORTREYLRVL 459
DB 545 KVLSSQFLNRWLIVCNRAHIEAPLSCLRRVGGSGMGNVGYPKLIIVDETPAAPFLRAI 604
QY 460 QOKVTADTQIVVCLLSSNRKDYDAIKKYLCTDCTPTSCQVARTLGKQOQVMAIATKIA 519
DB 605 QVHGDEPQVLMCILPSNQNYDSIKKYLSSDCPVPSQCVLTRLNKQGTMLSVATKIA 664
QY 520 LQWNCRMGELMRYDIPKLVMIVGIDCHDMTAGRRSIAGFVASINEGTRWFSRCIFQ 579
DB 665 MQMTCGLGELMSVEIPLSLMVGVDICRDLNKNVWVGVFASINSRITWFSRCVLIQ 724
QY 580 DRQOELVDGLKVCLOAALRAWNSCNEYPMSRIIVYRDGVDGQKTLVNYEVPQELDCLK 639
DB 725 RTIADADIADCLKVCMTGALNRWYHNDHLPARIYVYRDGVDGQKTLVNYEVPQELDCLK 784
QY 640 SIGRGNPRITVIVKRVNTRFFAOSGRLOPLPGTVIDVEVTRPEWYDFIVSQAVR 699
DB 785 ECG-----SDAR-----YDFYLSQATN 802
QY 700 SGQSVSPHYNVIYDNLGKPDHITQRLTYKLCIHYNNWPGVIRVPAPCOVAHKLAFVNGS 759
DB 803 RGTVSTHYNVIYDNLGKPDHITQRLTYKLCIHYNNWPGVIRVPAPCOVAHKLAFVNGS 862
QY 760 IHREPNSLSNRLLYL 775
DB 863 VHKEPSLELANNLFYL 878

```

RESULT 12

```

Q8N8G9 PRELIMINARY; PRT; 666 AA.
AC Q8N8G9;
DT 01-OCT-2002 (T8EMBLrel. 22, Created)
DT 01-OCT-2002 (T8EMBLrel. 22, Last sequence update)
DT 01-MAR-2003 (T8EMBLrel. 23, Last annotation update)
DE Hypothetical protein FLJ39518 (piwi).
OS Homo sapiens (human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N. A.
RA Tanigami A., Fujiwara T., Shibahara T., Goto Y., Hirao M., Shimizu F.,
RA Wakebe H., Ono T., Hishigaki H., Watanabe T., Ozaki K., Sugiyama T.,
RA Irie R., Otsuki T., Sato H., Wakamatsu A., Ishii S., Yamamoto J.,
RA Isono Y., Kawai-Hio Y., Saito K., Nishikawa T., Kimura K.,
RA Yamashita H., Matsuo K., Nakamura Y., Sekine M., Kikuchi H., Kanda K.,
RA Magatsuna M., Murakawa K., Kanekori K., Takahashi-Fujii A., Oshima A.,
RA Sugiyama A., Kawakami B., Suzuki Y., Sugano S., Nagahara K.,
RA Masuko Y., Nagai K., Isogai T.;
RT "NEDO human cDNA sequencing project."

```

RL Submitted (JUL-2002) to the EMBL/GenBank/DBJ databases.

DR EMBL: AK096337; BAC04873.1; --
 DR InterPro: IPR003100; PAZ.
 DR InterPro: IPR003165; Pwi.
 DR Pfam: PF02170; PAZ; 1.
 DR Pfam: PF02171; Pwi; 1.
 DR PROSITE: PS0821; PAZ; 1.
 DR PROSITE: PS0822; Pwi; 1.
 KW Hypothetical protein.
 SQ SEQUENCE 666 AA; 76099 MW; 1082D86916390FA8 CRC64;

Query Match 44.8%; Score 1838.5; DB 4; Length 666;
 Best Local Similarity 51.5%; Pred. No. 8.8e-138;
 Matches 346; Conservative 127; Mismatches 190; Indels 9; Gaps 4;

```

QY 107 ITITLNEIPPTSCLOFYNIIFRELLKIMNLQOIGRNYNPNPDIPIPSHRLVWPFG 166
DB 1 MTITLRELPSSPVCIOVFNIFRILKIMLSYQIGRNYFSEVEIP-----QY-NKLFNADV 60
QY 167 TTSILOVENSIMLCTDVSHKVLSETVLDPMFNFYHQTTEHKFQEQVSKELIGLVLTIKY 226
DB 61 AISVSYFERKLLFSADVSXKVLNSETVLEFMTALCOETGLSCFTQTCERKQLGLIVLTY 120
QY 227 NNKTYRVDIDWDQNPXSTFKKADGSEVSEFLEYYRYKYNQBEITDLKQPVLVGQPKRRR 286
DB 121 NNRTYIDDDWSVKPTHTFLKRDGTEITVYDYKQYDITLSDNQPVLSLKKRND 180
QY 287 GGTLPGLMILPELCVLTGLTDMRNDENOMKDLAVHRLTPQORQREYGRILDIYHKND 346
DB 181 NSB-PQALHILPELCFTGLTDOATSDFOIMKAVAEKTRLSFGQORLARLVDDIORT 239
QY 347 NVQRELRDGLSFDPSNLLSFGRILOTEKHOGKTFDYNPQ---FADWSKETRGAFLIS 403
DB 240 NARFELETWGLHFGSQ-ISLTGRIVSEKILMQ---DHICQPVSAADWSKDITCKILN 294
QY 404 VKPDLNWLIIYTRNVEAANSILQNLKVTYPAMGMQMKAMIEVDORTREYLRVLQKV 463
DB 295 AQSLNTMLICSDRTEYVAESFLNCLRRVGGSGMGNVGYPKLIIVDETPAAPFLRAI 354
QY 464 TADTQIVVCLLSSNRKDYDAIKKYLCTDCTPTSCQVARTLGKQOQVMAIATKIALQMN 523
DB 355 DPDVQVLMCILPSNQNYDSIKKYLSSDCPVPSQCVLTRLNKQGTMLSVATKIAMQNT 414
QY 524 CKMGELMRYDIPKLVMIVGIDCHDMTAGRRSIAGFVASINEGTRWFSRCIFQDRQ 583
DB 415 CKLGELMRYEIEPLKSLMVGVDICRDLNKNVWVGVFASINSRITWFSRCVLIQ 474
QY 584 ELVDGLKVCLOAALRAWNSCNEYPMSRIIVYRDGVDGQKTLVNYEVPQELDCLKSICR 643
DB 475 DVADCLKVPMTGALKNKYKYNHDLPARIYVYRDGVDGQKTLVNYEVPQELDCLKSICR 534
QY 644 GYNPRITVIVKRVNTRFFAOSGRLOPLPGTVIDVEVTRPEWYDFIVSQAVRSGSV 703
DB 535 NTSRLSIVIVRKKCMRPFETENNRVTQNPPLGTVVDDSEATRENEWYDFYLIISQVACRGTV 594
QY 704 SPHYNVIYDNLGKPDHITQRLTYKLCIHYNNWPGVIRVPAPCOVAHKLAFVNGSIRHE 763
DB 595 SPHYNVIYDNLGKPDHITQRLTYKLCIHYNNWPGVIRVPAPCOVAHKLAFVNGSIRHE 654
QY 764 PNLGSLNRLLYL 775
DB 655 PSLANLHLYL 666

```

RESULT 13

```

Q9GPA7 PRELIMINARY; PRT; 808 AA.
ID Q9GPA7
AC Q9GPA7;
DT 01-MAR-2001 (T8EMBLrel. 16, Created)
DT 01-MAR-2001 (T8EMBLrel. 16, Last sequence update)
DT 01-OCT-2002 (T8EMBLrel. 22, Last annotation update)
DE Seawi (Fragment).
OS Strongylocentrotus purpuratus (Purple sea urchin).

```

OC Eukaryota; Metazoa; Echinodermata; Eleutherozoa; Echinozoa;
OC Echinoidea; Euechinoidea; Echinacea; Echinoida; Strongylocentrotidae;
OC Strongylocentrotus.
OX NCBI_TaxID=7668;
RN [1]
RN SEQUENCE FROM N.A.
RP Rodriguez A.J., Bonder E.M.;
RT "Seawi - Cloning and Molecular Characterization of a Sea Urchin
RT Homolog of Pw1";
RL Submitted (NOV-2000) to the EMBL/GenBank/DBJ databases.
RL EMBL: AV014500; AAG42534.1; -
DR InterPro: IPR003100; PAZ;
DR InterPro: IPR003165; Pw1.
DR Pfam: PF02170; PAZ; 1.
DR Pfam: PF02171; Pw1; 1.
DR PROSITE: PS0821; PAZ; 1.
DR PROSITE: PS0822; Pw1; 1.
FT NON TER 808
SQ SEQUENCE 808 AA; 91366 MW; 19528F4B9D10474C CRC64;

Query Match 44.2%; Score 1811.5; DB 5; Length 808;
Best Local Similarity 49.4%; Pred. No. 1.7e-135;
Matches 354; Conservative 125; Mismatches 228; Indels 9; Gaps 6;

QY 14 VKESKSGSIIIVRLSTNFRSLTSPQWALYOVHIDYNPLMEARRLSALLQHEDLIGK 73
DB 98 VKQAIAGDK---IALLANGFKLTKEDWQYQYVDFEPEILNPRARFALLKGSALLGK 154

QY 74 CHAFDTEILPLKRLQKQVTEFSKTRNGEDVRITITLNLPPPTSCLOFYNIIFRL 133
DB 155 GLTLDMWTMSLFLKAEKVTNLSAERKDGSNVNICVTHVATLNPAPNTLHLNVLPRC 214

QY 134 LKIMNLQOIGRNNYNDPDIIPSHLVLVPGFTTSILOVENSIMLCTDVSHKVLRS 193
DB 215 LKMINNEQVGRNYDPTAIDIKQHLQWPGFVTSILOYVDVLLSDISHKVLR 274

QY 194 LDFMNFYHQTEBHKFQEQVSKELIGLAVLTLYNNKTYRVDIDMDQPKFTFKADGSE 253
DB 275 WEVNDLFENKA-RGRFEKEITKLMIQOVLVLYNNKTYRVDIDFDTPADTFERSG-P 332

QY 254 VSLEYRKYQNGEITDLKQPVLSOPKRRGGTLPAPMLIPELCYLGLTDMKEND 313
DB 333 VSVVDYFKSYERVIHVDVQMLVSRPKRREKGV--GPAFLPELCLTGLSDMEAD 390

QY 314 FNVKDLAVHTRLTPEQREVEGRLLIDYIHKNDNVQRELRLDGLSFDNLSFSGRILQT 373
DB 391 FNVKDMGQHTRVGPQDRCKTLTSGFIKLSNEEVKTYLDSWGMFDEQVKLTGRVLP 450

QY 374 EKHQGGKTFDYNPQADRSKTRGAPLISVKPLDNWLLIYTRNVEANSLIQNLFKVT 433
DB 451 EKLFORQKQSPNPSNADWSRDRGNALTDATLNNWKIFYTRRANRGQDFIKSLRVA 510

QY 434 PAMGMQMKAIMTEV--DDTEAYLRVLQKVTADTQIVVCLLSSNRKDKYDAIKYLCTD 492
DB 511 NPMGMVVRGPEIIVELPDRTETVTSIQAQDQIVVVILPTNEKORYDAIKTCVVT 570

QY 493 CPTPSQCVVARTLGKQOVTMAITKALQNMCKMGELMRVDIPLKLVIMVIGIDYHMT 552
DB 571 HPCPSQVIVSRTLSKQMLMSVATKIAMQNMCKMGDLNRVETPLSLNLIIGIDSYHDS 630

QY 553 AGRSIAGFVASINEGTRWFSRCIPQDRGOELVDGLKVCLOAALRAWNSCNEYPSRII 612
DB 631 TGRSVLGFVAVNMKQTSFSSCAFOHAQGEFGANLSTLMNALKRYIQINEKPERII 690

QY 613 VYRDGVDGQGLKTLVNYEVPQFDLCL-KSIGRGNPRLTVIVVKKRVNTRFFAQSGBLQ 671
DB 691 IFRDGVGDSQVNLVDVDELKQIKDLDKVPQGTVHKLAUVVKKRINNRFPANLRGGLS 750

QY 672 NPLPGTVIDVETRPWYDFEIVSQAVRSGSVSTHYNNVINDNSGLKPKPHIQRLTY 727
DB 751 NPPPGTVIDVETKPHLYDYFIISQSVRQSGSVSTSYNNVYDPTGLKPDHMQRLTY 806

RESULT 14
Q9JMB6 PRELIMINARY; PRT; 971 AA.
ID Q9JMB6
AC Q9JMB6;
DT 01-OCT-2000 (TREMELrel. 15, Created)
DT 01-OCT-2000 (TREMELrel. 15, Last sequence update)
DT 01-OCT-2000 (TREMELrel. 22, Last annotation update)
DE MILI (Miwi like).
GN PIWIL2 OR PIWIL1 OR MILI.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RN SEQUENCE FROM N.A.
RP Miyagawa S.K., Kimura T., Nakano T.;
RT "Molecular Cloning and characterization of Piwi family genes";
RT Submitted (SEP-1999) to the EMBL/GenBank/DBJ databases.
DR EMBL: AB032605; BAA93706.1; -
DR MGD; MGI:1930036; Piw1.2.
DR InterPro: IPR003100; PAZ;
DR InterPro: IPR003165; Pw1.
DR Pfam: PF02170; PAZ; 1.
DR Pfam: PF02171; Pw1; 1.
DR PROSITE: PS0821; PAZ; 1.
DR PROSITE: PS0822; Pw1; 1.
SQ SEQUENCE 971 AA; 109488 MW; 01E143C6513310FB CRC64;

Query Match 43.1%; Score 1766; DB 11; Length 971;
Best Local Similarity 42.9%; Pred. No. 9.5e-132;
Matches 327; Conservative 162; Mismatches 264; Indels 10; Gaps 9;

QY 18 KYGSSGIIIVRLSTNFRSLTSPQWALYOVHIDYNPLMEARRLSALLQHEDLIGKCHAF 77
DB 214 KQSKGTPOSGLNLKIQCHNE-AVYQHVTFSPVCKSMFGLKXKHQSDHOSVGNVAF 272

QY 78 DGTILPLRLQKQVTEFSKTRNGEDVRITITLNLPPPTSCLOFYNIIFRLKIM 137
DB 273 DGSILYLPVKLQVVLKSKORTDDBAISISIKLTKILEPCSDLCIPFYVNVFRRVVKLL 332

QY 138 NLQOIGRNNYNDPDIIPSHLVLVPGFTTSILOVENSIMLCTDVSHKVLRSVLDPM 197
DB 333 DMKLVGRNYDPTSAVVLQOHLQWPGYASIRTDGGLFLLADVSHKVTINDSVLDVM 392

QY 198 FNFYHTEBHKFQEQVSKELIGLAVLTLYNNKTYRVDIDMDQPKFTFKADGSEVSFL 257
DB 393 HAIYQNKESH-FQDECSKLLVGSIVITRYNNYTRIDVDVWNTPKDSFVMSDGEITPL 451

QY 258 EYTRKOYNOEITDLKQPVLSOP-KRRGPGGTLPGPAMLIPELCYLGLTDMKENDNV 316
DB 452 EYYSKNYGTIVKEDDQPLIHRPSPQRNNHGMLLKGEITLLPELSFMTGIPKMKKDFRA 511

QY 317 MKDLAVHTRLTPEQREVEGRLLIDYIHKNDNVQRELRLDGLSFDNLSFSGRILQTEKI 376
DB 512 MKDLTQOINLSFKQHHGALCECLLQISQNETASNELTRWGLSHKDVHKIEGELLPMERI 571

QY 377 HQGGKTFDYNPQADRSKE-TRGAPLISVKPLDNWLLIYTRNVEANSLIQNLFKVT 435
DB 572 NLNRTSF-VTSEDNLNVKVEVTRDASILTI-PMHFALFYPKAMQDQARELVNMLEKIA 629

QY 436 MGMQMKAIMTEV--DDTEAYLRVLQK--KVYADTQIVVCLLSSNRKDKYDAIKYLCTD 492
DB 630 IGRIISPPAWELKDDRIETIYRTQSLGVGEGKIQMVVCIITMGTRDDLYGAIKLCCVC 689

QY 493 CPTPSQCVVARTLGKQOVTMAITKALQNMCKMGELMRVDIPLKLVIMVIGIDYHMT 552
DB 690 SPVPSQVINVRTIGQPTLRSVAQKILLQNMCKLGELGWVDIPLKQLMVIGMDVYHDS 749

QY 553 AGRSIAGFVASINEGTRWFSRCIPQDRGOELVDGLKVCLOAALRAWNSCNEYPSRII 612
DB 750 RGRSVVGFVASINLTLTKWRSVVFQMPHQBIVDSLCLVGLSKLYKYEVNHLPEKIV 809

QY 613 VYRDGVDGQGLKTLVNYEVPQFDLCLKSIGRGNPRLTVIVVKKRVNTRFFAQSGBLQ 672

```

Db      810 VYRDGVSQGLKTVANYEIPQOKCFEAD-NYHPRMVVQVKISTNLYLAAPDHFVT 868
Qy      673 PLPGTVIDVEVTRPEWYDFIVSQAVRSGSVPTNYNNIYDNGSLKPDHIOQLTYKLCHI 732
Db      869 PSFGTVVDHTITSCSEWVDFYLLAHVROCGGIPTHYICVLNTPANLSPDHMQRLTFKLCHM 928
Qy      733 YNNWPGVIRVPAPCOYAHKLAFLVGOSIHREPNLSISNRLYYL 775
Db      929 YNNWPGTIRVPAPCKYAHKLAFLSGQILHHEPAIQLCGNLFFL 971

RESULT 15
Q8CDG1 PRELIMINARY; PRT; 971 AA.
AC Q8CDG1;
DT 01-MAR-2003 (TREMELrel. 23, Created)
DT 01-MAR-2003 (TREMELrel. 23, Last sequence update)
DT 01-MAR-2003 (TREMELrel. 23, Last annotation update)
DE Piwi like homolog 1.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
SEQUENCE FROM N.A.
RP STRAIN=C57BL/6J; TISSUE=Testis;
RC MEDLINE=22354683; PubMed=12466851;
RA The FANTOM Consortium,
RA the RIKEN Genome Exploration Research Group Phase I & II Team;
RT "Analysis of the mouse transcriptome based on functional annotation of
RT 60,770 full-length cDNAs.";
RL Nature 420:563-573 (2002).
DR ENBL; AK030116; BAC26791.1; -.
SQ SEQUENCE 971 AA; 109458 MW; B63FE11A2E9DA0F9 CRC64;

```

```

Query Match      43.0%; Score 1764; DB 11; Length 971;
Best Local Similarity 42.9%; Pred. No. 1.4e-131;
Matches 327; Conservative 162; Mismatches 264; Indels 10; Gaps 9;

Qy      18 KTSSGIIVRLSTNHPRLTSRPOWALYQYHIDYNPLMEARRLRALLFOHEDLIGKCHAF 77
Db      214 KQSKGTPTQSLGLNLKIQEHE-AVYQHVFPSPSVCKSMRFGMLKDHQSVGTGNVAF 272
Qy      78 DGTILFLPKLQKQVTEVSKTRANGEDVRITITLWELPPTSCLOFNIIFRLLKIM 137
Db      273 DGSILYLPKLVQWELKQSKRTDDAEISIKIQLTKILEPCSDLCIPFYNNVFRVMKLL 332
Qy      138 NLOQIGRNYNPNPDIDIPSHRLVWPFGTTISLOYENSIMLCTDVSHKVLRSSETVLDPM 197
Db      333 DMKLVGNFVDPTSAVVLQOHLQIWEFYAASIRRTDGGFLFLLADVSHKVINRDSVLDVM 392
Qy      198 FNYHOTEBHKFOEQVSKELIGLVLTKNYNNKTYRVDDIDWQNPKSTPKKADGSEVSL 257
Db      393 HAIYQQNKHE-FODECSKLLVGSIVITRYNNRYIRDDVDNKTDPKDSFVMSDGEITPL 451
Qy      258 EYVRKYVNOEITDLKOPVLVSOE-KRRRGGTLPGLPAMLIPELCVLTGLTDMKRDENV 316
Db      452 EYISKYGITVKEDDQPLLIHRPSEKQNNHGMKKEILLPELSPMTGIPKANKDFRA 511
Qy      317 MKDLAVHTRITPQOREVGRLLIDYIHKNDNVORELDMGLSFDNSNLSFSGRILOTEKI 376
Db      512 MKDLTQGINLSPKQHGALRECLLQRIQSNEAAGNELTRWGLSLHKDVHKIEGRLPMERI 571
Qy      377 HOGGKTFDYNPQADWSKE-TRGAPLISVUKPLDNWLLIYTRRYEAAANSLIONLKVTPA 435
Db      572 NLRNTSF-VTSEDLNVMVKEVTRDASILTI-PMHFALFYPKRAMDQARELVNMLEKAGP 629
Qy      436 MGHQMKRAIMIEV-DDRTEAYLRVLOQ--KVTADTQIVVCLLSSNKKDYDAIKKYLCTD 492
Db      630 IGMRISPPAWVELKDDRIETIYRTIQSLIGVEGKIQWVCIITMGTRDDLYGAIKKLCVQ 689
Qy      493 CPTEFSCQVARTLGKQOTVMAIATKIALQNMCKXGGELMRVVDIPLKLVNIVGIDCYHDMT 552

```

```

Db      690 SPVPSQVINVRIGQPTRLRSVAQKILLQNMCKLGGELWGVDIPLKQLMVGMDVYHDFS 749
Qy      553 AGRSTAGFVASINEGWTWFSRCIFQDRGQELVDGLKVCLOALRANNSCNEYMPSRIL 612
Db      750 RGMRSVVGFAVINLTFLTKWYGRVVPQMPHQIIVDSLKLCLVGSLSKKYEVVNHCLPEKIV 809
Qy      613 VYRDGVGDGQLKTLNNEYVPOFLDCLKISIGRGYNPRLTIVVVKRVNTRFFAQSGLRLQN 672
Db      810 VYRDGVSDGQLKTVANYEIPQOKCFEAD-NYHPRMVVQVKISTNLYLAAPDHFVT 868
Qy      673 PLPGTVIDVEVTRPEWYDFIVSQAVRSGSVPTNYNNIYDNGSLKPDHIOQLTYKLCHI 732
Db      869 PSFGTVVDHTITSCSEWVDFYLLAHVROCGGIPTHYICVLNTPANLSPDHMQRLTFKLCHM 928
Qy      733 YNNWPGVIRVPAPCOYAHKLAFLVGOSIHREPNLSISNRLYYL 775
Db      929 YNNWPGTIRVPAPCKYAHKLAFLSGQILHHEPAIQLCGNLFFL 971

```

Search completed: December 4, 2003, 19:15:34
Job time : 45 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: December 4, 2003, 19:12:29 ; Search time 21 Seconds
(without alignments)
3549.082 Million cell updates/sec

Title: US-10-043-774B-2

Perfect score: 4102
Sequence: 1 MIFGVNTRONLDHYKESKTG.....VGOSTHREPNLSLSNRLLYYL 775

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 283308 seqs, 96168682 residues

Total number of hits satisfying chosen parameters: 283308

```
Minimum DB seq length: 0
Maximum DB seq length: 2000000000
```

Post-processing:	Minimum Match	0%
	Maximum Match	100%
	Listing first	45 s

```
Database : PIR_76:*
1:  _pir1:*
2:  _pir2:*
3:  _pir3:*
4:  _pir4:*
```

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Query %			DB	ID	Description
	Score	Match	Length			
1	1295.5	31.6	824	2	T20351	hypothetical prote
2	1162	28.3	722	2	T30995	hypothetical prote
3	557.5	13.6	834	2	T41568	argonaute-like prote
4	508.5	12.4	1014	2	H86438	protein t19823.7 [
5	505.5	12.3	891	2	T32079	hypothetical prote
6	498.5	12.2	1000	2	T23291	hypothetical prote
7	495.5	12.1	997	2	A84678	hypothetical prote
8	492	12.0	988	2	T52134	Argonaute (AGO1)-1
9	488.5	11.9	930	2	A84668	Zwille protein (im
10	482	11.8	813	2	JC5569	Argonaute (AGO1)-1
11	434	10.6	887	2	T01113	translation initia
12	429.5	10.5	1032	2	T23164	hypothetical prote
13	429.5	10.5	1035	2	T23165	hypothetical prote
14	421	10.3	958	2	S41013	hypothetical prote
15	411	10.3	1040	2	D85668	protein ZK757.3 [i
16	384	9.4	898	2	T25164	hypothetical prote
17	343.5	8.4	1139	2	T33275	hypothetical prote
18	316.5	7.7	905	2	T23510	hypothetical prote
19	283	6.9	892	2	T15195	hypothetical prote
20	280.5	6.8	945	2	T23965	hypothetical prote
21	253	6.2	965	2	T29933	hypothetical prote
22	252	6.1	1057	2	T16676	hypothetical prote
23	248.5	6.1	1032	2	T19324	hypothetical prote
24	240	5.9	1030	2	T18114	hypothetical prote
25	238	5.8	990	2	F89525	protein T22H9.3 [i
26	230.5	5.6	697	2	T15179	hypothetical prote
27	225	5.5	939	2	T18974	hypothetical prote
28	223	5.4	324	2	D87774	protein C24A11.3 [
29	220	5.4	881	2	T31818	hypothetical prote

30	219.5	5.4	674	2	T19268	hypochemical prote
31	219	5.3	363	2	T34486	hypochemical prote
32	200	4.9	567	2	T34339	hypochemical prote
33	189.5	4.6	706	2	H70425	conserved hypotet
34	168	4.1	944	2	T18911	hypochemical prote
35	145	3.5	713	2	H64464	hypochemical prote
36	129	3.1	185	2	T27784	hypochemical prote
37	129	3.1	709	2	G96610	probable disease r
38	123.5	3.0	2035	2	AC0333	yersiniabactin bio
39	123.5	3.0	2041	2	T17439	peptide synthetase
40	120	2.9	1261	2	S75130	sensory transducti
41	118	2.9	1120	2	H71664	transcription-repa
42	114.5	2.8	474	2	A6491	type III restricti
43	114	2.8	2035	2	A48654	probable polyketid
44	113.5	2.8	560	2	C38604	poly(3-hydroxyalka
45	113	2.8	949	2	G84486	probable plasma me

ALIGNMENTS

RESULT 1
T20351
A:Accession: T20351
A:Map position: 1
A:Map coordinates: 118/3; 223/3; 304/3; 452/1; 573/3; 632/1; 711/2; 773/3
A:Species: *Caenorhabditis elegans*
C:Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_1999
C:Accession: T20351
R:Wilkinson, J
submitted to the EMBL Data Library, June 1996
A:Reference number: Z19261
A:Accession: T20351
A:Status: preliminary; translated from GE/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-824 <W1>
A:Cross-references: EMBL:Z73906; PIDN:CAA98113.1; GSPDB:GN00019; CESP:D2030.6
A:Experimental source: clone D2030
C:Genetics:
A:Gene: CESP:D2030.6

Query Match	31.8%;	Score	1295.5;	DB 2;	Length	824;	
Best Local Similarity	34.9%;	Pred. No.	6.5e-86;				
Matches	278;	Conservative	169;	Mismatches	302;	Indels	47;
Gaps	15;						

QY	16	ESKTSGGSIIVRLSTNHFL--TSRFQWALYOHVIDYNELMEARLRSLALLFOHE--DLIG	72
DBb	40	QSKTSSGGQPKCFANFIPIEMTQSDYSIYQHVFEFTVDSKANREKMLDNNVTDIG	99
QY	73	KCHAFDGTILFLPKRLQOKVTFVSKTRNGEDVRITITLNLNLPPTSPCTCLQFVNIIPRR	132
DBb	100	HHFVDFDMILYKKEEQNQMTIEVQHPIDRSLICIRPQTNRRFLVDDPQOTINPNTIIRR	159
QY	133	LLKIMNLQIGRNYNPNPDIPIDPSHRLVIMFGFTSILQYENSIMLCTDVSHKVLRSCT	192
DBb	160	SFDALQUTQIGRNYFNWGSRAVPDYNMSILPGYETALRMYYEENFMFLCENRPFKMWREES	219
QY	193	VLDPMFNFYH-----QTBEHFQEQVSKELIGLVLTKNYNNKTYRVDDIDWDQNPKSTF	246
DBb	220	----MYILPHKELRSQNNPQVRQEKMEYGGTIIIFRNNKLHRYTRLDYDYSISPLSEF	275
QY	247	KKADGSVSFLTYRRKYQNOEITDILKQVLVSQPKRRRPGCGTTPGGPAMLIPELCYLTGL	306
DBb	276	VK-DGOSITLKEYFKNQYGIETVDDQDPIIISSEKPKQ--PGEPPQVSYYTPELCFPTGL	332
QY	307	TDKMRNDFNWKDLAVHTRLTPEQQRQREVGRLIDYIHKNDNVQRELDRDGLSPDSNLLSF	366
DBb	333	TDENRKDFQWKKIAKHTNRSSQQRLVRSKLIIVDLSSNEKVMCEFKYIGLISGQDIANV	392
QY	367	SGRIQLQTEKHQGGKTFDYNPQFADWSKETRGAPLISVKPLDNNLLIYTRNVEA--ANS	424
DBb	393	QARVLKSEPT--QCKKT--VEGQOAWARQVKECGIYRGSNNMTNVIYTPGSSNGHLSOK	449

```

Qy 425 LIQNLFKVTPAMGMQMRKAIMIEV-----DDRTAYLRVLQKVADTQIVVCLLSSNRK 479
Db 450 FIEARRLGLKIQVLQGLGPMCPVIRGISPNLDYLEGVGAIRKQVGDGEDIHMLVWMLADNKK 509
Qy 480 DKYDAIKKYLCTDCPTPOCVVARTL-----GKQOTVMAATKIALQMNCKMGGLWR 532
Db 510 TRYDSLKFLVCCECPHQCNVLRTRLAGSKDGGENKLGSLVLKVLQMICTGGALWK 569
Qy 533 VDIPKLVMIVGIDCYHDMTAGRRSIAGFVASINEMGTWFSRCIFQDRGOELVDGLKVC 592
Db 570 VNIPLKNTMIYGVLDYHDLSTLKGTVGACVSTSDNFTQFYSQTRPHENPTQGNLTHF 629
Qy 593 LQAALRAMNSNEY-WPSRIIVYRQVGVGQKTLVYVYVQFLDCLKSIGR-----643
Db 630 VRKALKQYDSNDQFLPSRLILYRDGAGDGOIPYIKNTEVLRDADACDAVTKAAELSNK 689
Qy 644 -GYNPRLTVIVVKKVNRTRFAQSGRLQ---NPLPGVIVDEVTRPEWYDFEIVSQAVR 699
Db 690 VOEKIKLAFIIVTKVNMRLKQ-GSSLDNAINPQPGVVDVTRPERMDYILVFPQVFN 748
Qy 700 SGVSPTHYNVYNSGLKPDHIOQLTYKLCHYIYNWPGVIRVPAPCOYAHKLAFLVGQS 759
Db 749 QGTVPVSYNIIHDDTDLGPKHQQLAFKLCYLYYNWQTVRVPAPCOYAHKLAFLTAQS 808
Qy 760 IHRPNSLSNRLYYL 775
Db 809 LHDDANGCLRDLPFL 824

RESULT 2
T30995
hypoetical protein C01G5.2 - Caenorhabditis elegans
C:Species: Caenorhabditis elegans
C>Date: 22-Oct-1999 #sequence_revision 22-Oct-1999 #text_change 29-Oct-1999
C:Accession: T30995
R:Bradshaw, H.; Stellyes, L.
submitted to the EMBL Data Library, August 1999
A:Description: The sequence of C. elegans cosmid C01G5.
A:Reference number: Z20956
A:Accession: T30995
A>Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-722 <BRA>
A:Cross-references: EMBL:U50068; PIDN:AAB37734.1
A:Experimental source: strain Bristol N2
C:Genetics:
A:Map position: IV
A:Introns: 58/3; 121/3; 202/3; 350/1; 471/3; 530/1; 609/2; 671/3
A:Note: C01G5.2

Query Match 28.3%; Score 1162; DB 2; Length 722;
Best Local Similarity 33.6%; Pred. No. 2.7e-76;
Matches 256; Conservative 154; Mismatches 279; Indels 74; Gaps 14;

Qy 41 WALYQVHIDYNPLMEARLRSLALFOHEDL--IGKCHAFDGTILFLPKRLOOKTEVFSK 98
Db 6 YSIYQYHVEFPTVDSKATRENLRQPSVTVEIGKHFVFDGMILYLKEEWDQNMIEVQH 65
Qy 99 TRNGEDVRITITLNLPTSPFTCLQFVNIIFRLLKIMNQQIGRNYNENDPIDPSH 158
Db 66 PNDNSLICIRFKKINRFLVDDPQINFTNITIRSFAMKLTQLGRNVD-----115
Qy 159 RLVTWPQFTTSILOYSIMLCTDVSHKVLRSFTVDFMNFYHQTBEHFKQEQVSKELI 218
Db 116 -----WDNSRAL-----RKELRS-----CONNRQRVQEKQNEVYG 145
Qy 219 GLVLTLYKYNKTYVDDIDWDQNPFSFKKADGSEVSEFLYRQYQVQCEITDLKQPLVS 278
Db 146 GSTIITRNKHLRTRLDNEITLSFKQK-DGQIILKEFKYKQNDIDITDDQFIIS 204
Qy 279 QPKERRGGTLPQGFAMLIPELCYLTGLTQWRNDENVMDLAVHTRLTPEQRQREVGRL 338
Db 205 ESKPKQ--PGEPPQVIVPELCPFTGLTDMRKDFKQMKELAKHTRMSPQOQLDETRKL 262

```

```

Qy 339 IDYIHKNDNVQRELRDWGLSFDLSNLSFSGRILQTEKIHOQKTFDYNPOFADSKETRQ 398
Db 263 ITKLSQNOTMECCFQYWGISLQDLANQARVLKSEPL-QGKR--QYEGKQAEWARGVKE 319
Qy 399 APLISVPLONWLLIYTRRYEA--ANSLIQNLFKVTPAMGMQMRKAIMIEV-----DDR 451
Db 320 CGIYRGSNMTNWIIVPGSGNSGILLAQKFAEARNLGRITLQVLGEPWCVKINGISPNDY 379
Qy 452 TEAYLRVLQKVADTQIVVCLLSSNEKDKYDAIKKYLCTDCPTPOCVVARTL-----505
Db 380 LEGKAAIKVSDGEIHHMLVWMLADNKNTRYDSLKLYLCECPINQCVNLRITLAGSKD 439
Qy 506 -GKQOTVMAATKIALQMNCKMGGLWRVDIPLKVMIVGIDCYHDMTAGRRSIAGFVAS 564
Db 440 GGENKLGSLVLKVLQMICTGGALWKVNIPLKSTMIYGVLDYHDLSTLKGTVGACVST 499
Qy 565 INEGMTWFSRCIFQDRGOELVDGLKVCLOAALRAMNSCNE-YMPSRIIVYRQVGVGQQL 623
Db 500 TSNDFTQFYSQTRPHENPTQGNLTHFVRKSLKQYVYDNDKTLPSRLILYRDGAGDGOI 559
Qy 624 KTLVNVYVQFLDCLKSIGR-----CYNPRLTVIVVKKVNRTRFAQ--SGGRLQN 672
Db 560 PYIKNTEVLRDADACDAVTKAAELSNKVOEKIKLAFIIVTKRNMRLKQSGSSKSGAIN 619
Qy 673 PLPCTVIDEVTRPEWYDFEIVSQAVRSGVSPHYNVYDNLGKLPDHIQRLTYKLCHI 732
Db 620 PQPGTVDDTVTRPERMDYILVFPQVFNQGTVPVSYNIIHDDTGLGDPKHQQLAFKLC 722
Qy 733 YYNWPGVIRVPAPCOYAHKLAFLVGQSIHBPNSLSNRLYYL 775
Db 680 YYNWQTVRVPAPCOYAHKLAFLTAQSLHDDANGCLRDLPFL 722

RESULT 3
T41568
argonaute-like protein - fission yeast (Schizosaccharomyces pombe)
C:Species: Schizosaccharomyces pombe
C>Date: 03-Dec-1999 #sequence_revision 03-Dec-1999 #text_change 21-Jan-2000
C:Accession: T41568
R:Wood, V.; Rajandream, M.A.; Barrell, B.G.; Murphy, L.; Harris, D.
submitted to the EMBL Data Library, May 1998
A:Reference number: Z21391
A:Accession: T41568
A>Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-834 <WOO>
A:Cross-references: EMBL:AL023705; PIDN:CAAL19275.1; GSPDB:GN00066; SPDB:SPCC736.11
A:Experimental source: strain 972h-; cosmid c736
C:Genetics:
A:Gene: SPDB:SPCC736.11
A:Map position: 3
A:Introns: 43/3
C:Superfamily: rabbit translation initiation factor eIF-2C

Query Match 13.6%; Score 557.5; DB 2; Length 834;
Best Local Similarity 22.9%; Pred. No. 3e-32;
Matches 198; Conservative 155; Mismatches 346; Indels 167; Gaps 31;

Qy 20 GSSGIIVRLSTNHFRTSPQWALYOHIDYNPLMEARLRSLALFOHEDL-----70
Db 16 GGLGKQITLKANFPQIISLFNETINQYHIVGVDSRVPKRSQLIWNSEKVEKYFGSSWM 75
Qy 71 -----IGKCHAFDGTILF-----LPKELQ--QKVTEV-----PSKTINGED 104
Db 76 NSYVDGRSMCWSGDADGTIKVNIIGSESHPREIFSIQKSKINLHTLSQFVNSKYSSD 135
Qy 105 VRITITLNLPTSPFTCLQFVNIIFRR-----LLKIMNQQIGRNYNENDPIDPSH 159
Db 136 -----PQVLSSTMFLLDLKPKSETLFGFMGSPFTGNGVSLGGVGE-----178
Qy 160 LVTWPQFTTSILOYSIMLCTDVSHKVL-RSETVLDFM-----197

```

Db 179 --AWKGFVQSIQRENGFMSVNVDISSAFWRNDSLLQILMEYTDGNSVRDITRFDKELS 236
Qy 198 --FNFVHQTEHFKFQEQSKELIGLVLTKNKTYRVDIDWDQNPKEST-KKADQSE- 253
Db 237 KRFRFLKVTQHRNN-----VGTDLANRVYSIEGSSKASQSFVRRLNGEEQ 285
Qy 254 --VSFLFYRKOYNQIBITDLQPVLYSQPKRRRGGTILPGFAMLIPELCYLT--GLTDK 309
Db 286 KISVAEYPLENHNVRLOYNLPCIL-----VANGAMLPFIECFVVGQRYTAK 333
Qy 310 MRDNFNVMDLAVHRTLPQEQREVRGLIDYIHKND-NVQBELRDWGLSPDSNLSFSG 368
Db 334 LNSD-----OTANMRFAVORFERVQOIDDVHQMDWDITDYLQYGMKIOKMWLEVA 388
Qy 369 RILOTEKHQGGKTDFYNQFADWSKETRGAPLISV--KPLDNWLLI---YTRR-NYBAA 422
Db 389 RYLETPSIRYGGDIE-RPVSGRW--LEGKFLPPPRAPIRSMVAMCTSTRRLPWRGI 445
Qy 423 NSLIQNLKVTTPAMQHO--MRKAIMIEVDNR--TEAYLRVLOQVY----TADTQIVVCLL 474
Db 446 ENFLQTYVQTLTSLGINFYMKKPPVLYADIRGSVEELCITLYKKAEOQVGNAPPDYLFFIL 505
Qy 475 SNRDKDKYDAIKYLCCTDPTSCQVARTLQKQOTVMAIAKIALQMNCKMG---ELW 531
Db 506 DKNSPEPYGSIKVCNMTLGVPSQCAISKHI--LQSKPYCANLGMKINVKVGGINCSLI 563
Qy 532 RVDIPLKV--MIVGIDCVH-DMTAGRRSIAGFVASINEGMRWFSRCIFQDRGOELVDG 588
Db 564 PKSNPLGNVPTILGDDVHPGVGATGVSIASVADLNGCKYTAVERSPRHOEVIEG 623
Qy 589 LKVCQAALRAWNSCNEYMPSRIIVRQGVGGQKLTLYNVVQFOLDKLSIGRGNPR 648
Db 624 MDIVVYLQGRFAMTKQOPRIIFRDTSGQFLSVINDLSQIKACHSLSPKYNPK 683
Qy 649 LTVIVVKKRWTRFFAQ--GGRQNLPLGTVIDVETRPWYDFFIQVQAVRSGSVSP 705
Db 684 ILVCTQKSHARFFIKNSDGRNGNPLPGTIEKHVTHPYOYDFYLSHPSLQGVSP 743
Qy 706 THYNYIDNSGLKPDHIORLYTKLCHYYNWGVTRVAPCOYAKLAPL----- 755
Db 744 VHYTVLHBIQPPDFQFQTLNCLVYARATSAVSLVPFVYIAHLVSNLARYQDVTADD 803
Qy 756 -----VGQSIHREPNSLSNRL 772
Db 804 TFVETSEASWDQEV--KPLALSSKL 827

RESULT 4

H86438
protein T19E23.7 [imported] - Arabidopsis thaliana
C:Species: Arabidopsis thaliana (mouse-ear cress)
C:Date: 02-Mar-2001 #sequence_revision 02-Mar-2001 #text_change 31-Mar-2001
C:Accession: H86438
R:theologos, A.; Ecker, J.R.; Palm, C.J.; Federspiel, N.A.; Kaul, S.; White, O.; Alonso,
Chain, C.W.; Chung, M.K.; Conn, L.; Conway, A.B.; Creasy, T.H.; Dewar, K.;
ansen, N.F.; Hughes, B.; Huizar, L.
Nature 408, 816-820, 2000
A:Authors: Hunter, J.L.; Jenkins, J.; Johnson-Hopson, C.; Khan, S.; Khaykin, E.; Kim, C.
C.A.; Li, J.H.; Li, Y.; Lin, X.; Liu, S.X.; Liu, Z.A.; Luros, J.S.; Maiti, R.; Marziali,
Rizzo, M.; Rooney, T.; Rowley, D.; Sakano, H.
A:Authors: Salzberg, S.L.; Schwartz, J.R.; Shinn, P.; Southwick, A.M.; Sun, H.; Tallon,
ker, M.; Wu, D.; Yu, G.; Fraser, C.M.; Venter, J.C.; Davis, R.W.
A:Title: Sequence and analysis of chromosome 1 of the plant Arabidopsis.
A:Reference number: A86141; MUID:21016719; PMID:11130712
A:Accession: H86438
A>Status: preliminary
A:Molecule type: DNA
A:Residues: 1-1014 <STO>
A:Cross-references: GB:AE005172; NID:G6692120; PIDN:AAF24585.1; GSPDB:GN00141
C:Genetics:
A:Gene: T19E23.7
A:Map position: 1

RESULT 5

T32079
hypothetical protein T07D3.7 - Caenorhabditis elegans
C:Species: Caenorhabditis elegans
C:Date: 29-Oct-1999 #sequence_revision 29-Oct-1999 #text_change 21-Jan-2000
C:Accession: T32079
R:Lamar, B.; Wamsley, P.; Twyman, B.
submitted to the EMBL Data Library, July 1997

Query Match 12.4%; Score 508.5; DB 2; Length 1014;
Best Local Similarity 23.9%; Pred. No. 1.5e-28;
Matches 209; Conservative 151; Mismatches 345; Indels 171; Gaps 40;
Qy 5 VNTRQNLHVKSCTGSSGII-----VRLSTNFRITSPOWALYQYHIDYN---PLMEAR 57
Db 153 VATSDSKBEMKRPDR--GVAVRRVNLVYVHYKFNFPESVIRHYRVEIKGEIKTKKVS 210
Qy 58 RLRSALLFOH-----EDLIGKCHAPDG-----TILFLPKLQKQKVFVSKTRNGEDVR 106
Db 211 RPELAVRDKVTDNDPDEPLAMTAYDQKNIFSAVELETG-SYKVEYKTEEMRGRSTV 269
Qy 107 ITITLNLPPPTSCLOQYNIIFRLKIMLQO--IGRNYNNDPI-----DIP 156
Db 270 FTIKQVN-----VLKGLKEYMTGRSSFNPRDVLQGMVVMKEHP 310
Qy 157 S-----HRLVWPFTTSLQYENSIMLCTDVSHKVLK--SETV 193
Db 311 SKCMITVGSFFTRTEPEDDFGVIAAGKYRHILKPTAQGLSLCLDYSLVAFKMSV 370
Qy 194 LDFMFPYHQTBEHKFOEO-VSKELIGLVV-----LTKNNKTYRVDID 237
Db 371 IEYLKLYFMSMDRQPRRDVBEELIGLVTVNHRKQKLAIVGLSMQMTKDIFDLID 430
Qy 238 WQDN--PKSTFKKADGSEVSLEYRKYNOETDLQPVLYSQPKRRRGGTILPGPM 295
Db 431 QEGNEPPRKT-----SIVEYFRIKYGRHIVHKDIFCLDLGKNGRN----- 471
Qy 296 LIP-ELCYLTGLTDKVRNDFNVKMLAVHTR---LTPEQOREVGRGLIDYIHKNDNVOR 350
Db 472 FVPMFCDL--VEGQIYPKNDLKDLSALMLKSLVNPQORQNRIDKMK-ARNGPSGGE 528
Qy 351 ELRDWGLSPDSNLSFSGILQTE--KIQGGKTFDYNP---QFADWSKETRGAPLISVK 405
Db 529 IICNFGKVDNTNTPVEGRVLKAPSLKLAERGVREEPNPRQNNQNMMLKKGVTGSI- 587
Qy 406 PLDNWLLI---YTRRNYEAANSILOLFKVTTPAMGMQMKAKIMIEVD-----DTEA 454
Db 588 -VKHVAVLDTASERFNKMPNDFVDNLIDRCWELGQMEAPIVYKTSRMETLSNGNAIEE 646
Qy 455 YLRVLOQVKT-----ADTQIVVCLLSNRKDKYDAIKKYLCCTDCTPPSQVVA--RTLQK 507
Db 647 LLRSVIDEASRKHGGRPTILVCLAM-SRKDDGKYTLKWAETKGLGLVTCFLTPATKGG 705
Qy 508 QOTVMAIATKIALQMNCKMG---ELWRVDIPLK---VMVIGDCVHDMTAGRR--SIA 559
Db 706 DQ-----YRANLAKMAKAKVGGSNVELMDTFSPFKKDEDEHFIAGVNHFAARDKMSPSIV 761
Qy 560 GFVASIN-EGMTRWFRGRCIFQDRGOELVDGL-KVCLQALRAWNSCNEYMPSRIIVYRDG 617
Db 762 AVGTLLNWPANRYAARVIAQPHRKEEIQGFGDACLE-LVKAHVQATGKRPKNKIVIFRDG 820
Qy 618 VGDQKLTLYNVVQFOLDKLSIGR-GYNPLRTVIVVKKRVNTRFFA---QSGGRLQNP 673
Db 821 VSDAQFMDNLNVEL---LDVQLTFKKNYKPKLTIVIAQKHQTRFRFPATNNDGSKGNV 877
Qy 674 LPGTVIDVETREWDYDFFTVSQAVRSGSVSPHYNYVINDNSGLKPDHIORLYTKLCHY 733
Db 878 PSGTVVDTKVIHPEYVDFYLSHHGGIGTSKPTHYVTLNDELGTSDQVQKLIFFECFTF 937
Qy 734 YNPFVIRVPAPCOYAKLAPLVGQSIH-----RPN 765
Db 938 TRCTKPSLVPPVYADWAFR-GRMYHEASSREKN 972

A:Description: The sequence of C. elegans cosmid T07D3.

A:Reference number: Z21121

A:Accession: T32079

A:Status: preliminary; translated from GB/EMBL/DBJ

A:Molecule type: DNA

A:Residues: 1-891 <LAM>

A:Cross-references: EMBL:AF016682; PIDN:AA66187.1; GSPDB:GN00020; CESP:T07D3.7

A:Experimental source: strain Bristol N2; clone T07D3

C:Genetics:

A:Gene: CESP:T07D3.7

A:Map position: 2

A:Introns: 34/2; 92/2; 206/1; 560/1; 879/3

C:Superfamily: rabbit translation initiation factor eIF-2C

Query Match 12.3%; Score 505.5; DB 2; Length 891;

Best Local Similarity 23.7%; Pred. No. 2e-28;

Matches 201; Conservative 139; Mismatches 338; Indels 169; Gaps 32;

QY 20 GSSGIIVRLSTNHRLTSRPMALYQYHIDYNPLMEARLRALLFQHEDLIGKCHAFDG 79

DB 52 GVEGRSILLRANFAVRI-PGGSVQHYQIDVPPDKCPRVNEVI-----G 96

QY 80 TILFLPKRLQCKVTEFSTRN-----GEDVRIITITNE-----LPTS 119

DB 97 CLISFSKYFTNIRPVYDGKRNMYTRPLIGTEPMNPEVTLPGDSAVERKFSVTMKWIG 156

QY 120 PTCLE-----QFYNIIFRLLKIMLQOIGRNYNPNDIPIDPSHRL 160

DB 157 QVCLSDALDAMEGRVQVPEHVAQSDIVLRH-LPSLKTYTPVGRSFFTPPGVMK-PGMQM 214

QY 161 -----VIWPGTTSILOVENSIMLCTDVSHKVL-RSETVLDFMNFVH----- 202

DB 215 HQBSKLGGRGVWFGEFGHQSVPQWMLNIDVSATAFYRAMEVIEFVAEVLLELPVQALA 274

QY 203 -----QTEHKEFQEQVSKELIGLVLTXY-----NNKTYRVDIDWDQPKSF-----KKA 249

DB 275 ERRLSDAQVKF-----THEIRGLKIEITHCGAVRRKRYCNVTRRPAQTQTFPLQLETG 330

QY 250 DGSEVFLKYRQYQNEITDLKOPVL-----VSQPKRRRGGPGTLPGPAMLIPELC----- 301

DB 331 QTEICTVAKYFPDKYR-----IQLVKPHLCLOVGEQKH-----TYLPEVCDIVP 377

QY 302 ---YLGLTDKNDNFNMKDLAVHRLTPEQOREVGRGLDIYHKNDVNQREL----- 352

DB 378 GQRLKLLTDVQSTW-----IKATARSAPF-RERBICKLVS-----KAELSADPPA 423

QY 353 RDWGLSFDNLLSFGRILOTEKIHOGG-KTFDYNPQFADWSKETRGAPLISVKPLDNW 410

DB 424 HEFGIITNPAMTEVKGRLVSAKLLYGGRRHATTALPNQGVW---DMRGKQFHTGMEVRTW 481

QY 411 -LLIYTRNVEAANSL-----IONLFKVTTPANGMOM-----RKAMIEVDDTEAYLRVL 459

DB 482 AIACFAQOQHVKENDLRFMTTQLORISTDAGMPIIGTTPMFCYASGVE---QVEPMFKYL 538

QY 460 QQKVADTQVWCLSSNRKDKYDALIKKYLCTDCTPTSCQVARTLKGQOOWMATATKIA 519

DB 539 KQYYSALQILVVVL---PGKTPVIAEVKRVGDTVLGIATQCVQAK---NAITPFIQLSNLC 594

QY 520 LQNKCRKGG-----ELWRVDIPLKLMIVGIDICVHDMTAGRR---SIAGFVASINEGMTR 571

DB 595 LKMNVLKGGVNSILLPNVRPIEFNEFVIFLGCDITHFAAGDTRKPSIAAIVGSGMDAHPSR 654

QY 572 WFSRCIFQDRGQELVDGLKYLQALRAWNSCNEYMPSRILVYVDCGVGQOLKTLNNVEV 631

DB 655 YAFVRVQQRQELITDITMVELLIVQFYRTRFRPARIVTRDGVSGQFLNVLQYEL 714

QY 632 PQFLDCLKISGRGNPLTIVIVKRVNTRFF-----AQSGRLQNPGLPTGVIDVEVTRPE 687

DB 715 RAIRACVWLESQVQPGITFIAVQKRHHTRLPAADKADQVKAFNIPPGTVDVGTIHTPT 774

QY 688 WYDFEIVSQARSGSVSPHTNYINYSGLKPHIQRLTYKLCHIIYNNWPGVIRVPAPCQ 747

DB 775 EEDFFLCSHAGIQGTSRPSHYLWDDNDITADELOQLTYQMCHTYVRCTRVSIPAPAY 834

QY 748 YAHKLAP 754

DB 835 YAHLVAF 841

RESULT 6

T22391

hypothetical protein F48F7.1 - Caenorhabditis elegans

C:Species: Caenorhabditis elegans

C>Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 04-Mar-2000

C:Accession: T22391; T23234

R:Coles, L.

submitted to the EMBL Data Library, February 1996

A:Reference number: Z19559

A:Accession: T22391

A:Status: preliminary; translated from GB/EMBL/DBJ

A:Molecule type: DNA

A:Residues: 1-1000 <WIL>

A:Cross-references: EMBL:Z69661; PIDN:CAA93496.1; GSPDB:GN00028; CESP:F48F7.1

A:Experimental source: clone F48F7

R:Coles, L.

submitted to the EMBL Data Library, February 1996

A:Reference number: Z19712

A:Accession: T23234

A:Status: preliminary; translated from GB/EMBL/DBJ

A:Molecule type: DNA

A:Residues: 1-1000 <WIL>

A:Cross-references: EMBL:Z69663; PIDN:CAA93512.1; GSPDB:GN00028; CESP:F48F7.1

A:Experimental source: Clone K02B9

C:Genetics:

A:Gene: CESP:F48F7.1

A:Map position: X

A:Introns: 70/3; 128/2; 185/2; 673/1

C:Superfamily: rabbit translation initiation factor eIF-2C

Query Match 12.2%; Score 498.5; DB 2; Length 1000;

Best Local Similarity 22.5%; Pred. No. 7.8e-28;

Matches 199; Conservative 139; Mismatches 322; Indels 225; Gaps 34;

QY 20 GSSGIIVRLSTNHRLTSRPMALYQYHIDYNPLMEARLRALLFQHEDLIGKCHAFDG 79

DB 145 GVEGRSILLRANFAVRI-PGCTIQHYQVDVTP-----DKC----- 179

QY 80 TILFLPKRLQCKV-----TEVSKTRNGEDVRITITLTNELP-----PTSP-- 120

DB 180 -----PRRVNREIISCLISAFSKYFTNIRPVYDGKRNMYTRRPLIGRBRMDPDLPGD 234

QY 121 -----TCLOFYNIIFRLLKIMLQOIGRN 145

DB 235 SAVERQPSVSLKWKVQVSLSTLEDAMEGRVQVYPPFAVQAMDVILRH-LPSLKTYFVGS 293

QY 146 YANPNDFIDIPSHRLV-----TWPGTTSILOY 173

DB 294 FFSF--PVPNASGVWAGSCPPQASGAVAGSAHAGSAGVHAESKLGGRGVWFGPHQSVRPS 351

QY 174 ENSIMLCTDVSHKVL-RSETVLDFMNFVH-----QTEHKEFQEQVSKELIG 219

DB 352 QWRKMLNIDVSATAFYRAMEVIEFIAEVLLELPVQALAEARLSDAQRVFP-----TKEIRG 407

QY 220 LVVLTKY---NNKTYRVDIDWDQNPKSTF---KKADGSEVSLFLEYRKYQYQNEITDLK 272

DB 408 LKIEITHCGQMRKRYCNVTRRPAQTQTFPLQLEGTQIECTVAKYFYDKYR---IQLK 464

QY 273 QPVL-----VSQPKRRRGGTLPGPAMLIPELCYLT-----GLTDKRNDFNVKMDLA 321

DB 465 YPHLPCLQVGEQKH-----TYLPEVCNIVPGQRCIKKLTVDVSTW-----IK 509

QY 322 VHTFELTPEQOREVGRGLI-----DYHKNDNVQRELDRWGLSFDNLLSFGRILO 373

DB 510 ATARSAPF-RERBISNLVRAKFAFSADPPAH-----EFGITINPAMTEVKGVLSA 558

QY 374 EKIHQGGKT-FDYNPQFADWSKETRGAPLISVKPLDNW-LLIYTRNVEAANSL-----IQN 428

Db 559 PKLLYGRTRATAPALNQGVM--DMRGKQFHTGIDVRVWAIACFAQQOQHVKENDLRMTNQ 616
Qy 429 LFKVTPAMQMOM-----RKAIMLEVDRTAEYLRVLQOKVTADTQIVVCLLSNKKDK 481
Db 617 LQRIENDAGMPIVGNPCCKAVGVE--QVEFMPFKYLKQNSYG-IQLVVVLV-PGKTPV 671
Qy 482 YDAIKKYLCTDCTPTSCQCVARTLKQQTVAIAIKALQNNCKMG--ELNRVDI 535
Db 672 YAEVXRVGDTVLGIATOCVQAK--NAIRTPQTLNCLKNVKNLGGVNSILLNVPRI 729
Qy 536 PKLVMIVGIDYCHDMTAGRR--STAGFVASINEGRTWFRSCIPQDRGQLVCLKVL 593
Db 730 FNEPVIFFGCDITHPAGDSRKSFAAAYVGSMDAHPRYAATVRVQQRQBIISDLTVMV 789
Qy 594 QAAALRAWNSCNMPSRIIVYDVGDKLTLVNVYVPPQFLDCLKSTGRVGNRLTIV 653
Db 790 RELVQFYNTFKFARIIVYDVGSEGFVNLQYELRAIEACMMLERGYPGIFTIA 849
Qy 654 VKKRVNTRFFA---QSGRLQNPUGTVIDVEVTRPWEYDFIVSQVRSVSPHYN 709
Db 850 VQKRHHTRFLFAVDKQDVQKAYNIPPGTVDVGIHTPTFEFDYLCGHAGIQTSRPSHYH 909
Qy 710 VYDYSGLKPDHQRILYKLYCHYYNMPGVIRVPAPCOYAHKIAF 754
Db 910 VLWDDNLTADLQQLTYQWCHTYVRCRSVSIPAPAYIAHLVAF 954

RESULT 7
A84678
Argonaute (AGO1)-like protein [imported] - Arabidopsis thaliana
C:Species: Arabidopsis thaliana (mouse-ear cress)
C:Date: 02-Feb-2001 #sequence_revision 02-Feb-2001 #text_change 16-Feb-2001
C:Accession: A84678
R:Lin, X.; Kaul, S.; Rounsley, S.D.; Shea, T.P.; Benito, M.I.; Town, C.D.; Fujii, C.Y.;
M.; Koo, H.; Moffat, K.S.; Cronin, L.A.; Shen, M.; VanAken, S.E.; Umayam, L.; Tallon, L.;
Eus, D.; Nierman, W.C.; White, O.; Eisen, J.A.; Salzberg, S.L.; Fraser, C.M.; Venter, J.
Nature 402, 761-768, 1999
A:Title: Sequence and analysis of chromosome 2 of the plant Arabidopsis thaliana.
A:Reference number: A84420; MUID:20083487; PMID:10617197
A:Accession: A84678
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-997 <STO>
A:Cross-references: GB:AE002093; NID:G4510428; PIDN:AAD21514.1; GSPDB:GN00139
C:Genetics:
A:Gene: At2g27880
A:Map position: 2
C:Superfamily: rabbit translation initiation factor eIF-2C

Query Match 12.1%; Score 495.5; DB 2; Length 997;
Best Local Similarity 24.1%; Pred. No. 1.3e-27;
Matches 200; Conservative 139; Mismatches 363; Indels 129; Gaps 30;
Qy 20 GSSGIIVRLSTNHFRITSPQWALYQYHIDYNPLMEA---RLRSALLFOHED--LIGK 73
Db 154 GTLGKVMVRANHF-LVQVADRLYHYDVSINPEVISTKVNVMKLLVKNVKSGLGK 212
Qy 74 CHAFD-----GTILFLPKLQOKVTEVFSTRNGEDVRITILNELPPTSTPCLOF 125
Db 213 SPAYDGRKSLYTAGLPFFDSKEFVNLAERKADGGSGKDRPFKAVKN---VTSTLYQL 269
Qy 126 YNIIFFR-----LKKIMNL-----QOIGRNYNPNPDIPSHR-----LVI 162
Db 270 QQFLDRKQREAPYDTTQVLVDVLRDPSNDYVSGSFHTSLGKDARGELGDGLEY 329
Qy 163 WPGFTTSILQYNSIMLCTDVSHKVLRSETVL--DFMNFYHOTEHK-----FQSQVSK 216
Db 330 WRGVFQSLRTOGLSLNIDVSARSFVEPIVTVDTFISKFINRLNRLRDSRLKVKV 389
Qy 217 LIGL-VVLTKYN-NKTYRVDIDWDQNPSTFKKADGSEVSFLEYKQYNOEITLQOP 274
Db 390 LRLTKVLLHWNCTKSAKISGLSIFIRELURFTLEDKSEKTVVQYFAEKYINRVKQALP 449

Qy 275 VLVSQPKRRRGCGTLPGPAMLIPELC-----YLTLGLTDQRNDFNVMKOLAVHTRILT 327
Db 450 AIQT-----GSDTRPVYLPMLCOIDEGQRYTKRLNEKO-----VTALLKATQOR 494
Qy 328 PQORQREVRLDYIHKNDNVORELDRWGLSPDNLSPSGLILQTE--KLHQGSKTPTY 385
Db 495 PPDRENSIKNLVVKNNYNDLSKE---FGMSYTTQLASIEARVLPPMLKYHDSKERNV 551
Qy 386 NPQFADWS---KETRGAPLISVKPLDNWLLI--VTRRNYEAANSLIQNLKFKVTPAMNQ 439
Db 552 NPLGQWAMIDKKNVNGAKVTS-----WTCVSFSTRIDRGLPQEFCKQLIGMCYKGM 605
Qy 440 MRKA-----IMIEVDDRTAYLRVIOQKVATDTQIVVCLLSNRKDKYDAIKKYLCTDCP 494
Db 606 FKQPAIPFISCPPEHIEBALDI--HKRAPGLQLLIVL-PDVTGSGYKIKRICETELG 662
Qy 495 TBSQCVVARTLKGQQTVAIAIKALQNNCKMG--ELNRVDIPL---KLVMIVGI 545
Db 663 TVSQCCQPRQVKNLN--KOYMNVALKINVKTGRTVNLDAIRNIPITDRPTIIMGA 720
Qy 546 DCVHDMTA--GRRSIAGFVASIN--EGMTRWFRSCIPQDRGQLVCLKQAAALR--- 598
Db 721 DVTHPQPGEDSPSIAAVVASMWPENIKYKGLVSAQAHRBIIQDLVLYLQDPQKGLWH 780
Qy 599 -----ANNSCNEYMESRIIVYDVGDKLTLVNVYVPPQFLDCLKSGRYNPEL 649
Db 781 SGLIREHFIAFRATQOIFORIIFYRDGVSEGSQVLLHEMTAIRKACNSIQENIVPRV 840
Qy 650 TVIWKRVNTRFFAOSGG-----RLQNLPLGTVIDVEVTRPWEYDFIVSQVRSVSV 703
Db 841 TFVIVQKRHHTRFLFPQHGNDRTDKSGNIQGTVDVTKICHENEDFYLNHSHAGQGS 900
Qy 704 SPHYNIVYDNLKPKDHPHQRILYKLYCHYYNMPGVIRVPAPCOYAHKIAF 754
Db 901 RPAHYHLLDENGFTADQLQMLTNLNCYTYARCTKSVISVIPPAYIAHLAALF 951

RESULT 8
T52134
Zwille protein [imported] - Arabidopsis thaliana
C:Species: Arabidopsis thaliana (mouse-ear cress)
C:Date: 20-Oct-2000 #sequence_revision 20-Oct-2000 #text_change 03-Nov-2000
C:Accession: T52134
R:Laux, T.
Submitted to the EMBL Data Library, January 1998
A:Reference number: Z25975
A:Accession: T52134
A:Status: preliminary; translated from GB/EMBL/DDBJ
A:Molecule type: mRNA
A:Residues: 1-988 <LAU>
A:Cross-references: EMBL:AJ223508; PIDN:CAAL1429.1
C:Genetics:
A:Gene: zwille
A:Map position: 5
C:Superfamily: rabbit translation initiation factor eIF-2C

Query Match 12.0%; Score 492; DB 2; Length 988;
Best Local Similarity 23.6%; Pred. No. 2.3e-27;
Matches 204; Conservative 133; Mismatches 397; Indels 132; Gaps 32;
Qy 5 VNTQNLDHVKESKTSGGIIVRLSTNHFRITSPQWALYQYHIDYNPLMEAARLSAL- 63
Db 120 MGVRKNSNFAPRGFGTGLTKCIVKANHF-LADLPKDLNQLQYDVTITPEVSSKVNRAII 178
Qy 64 -----LFQHEDLICKHAFD-----GTILFLPKLQOKVTE---VFSTRNGEDVR 106
Db 179 AELVRLYKESDLGRPLPAYDGRKSLYTAGELPFTWKEFSVKIVDEDDGIINGKRERSYK 238
Qy 107 ITIT-----LTNELPPTSTPCLOFNIIFRLLKIMNLQOIGRNYNPNNDPI 153
Db 239 VAIKFVARANMHLGSLFGLAKRADCPQEAQVQIILVLR-LSVKRFPVGRSFPSPD--I 295

```

154 DIP--SHRLVWPQFTSIILOYENSIMLCITDV-SHKVLRSETVLDFMNFYQTEEHK- 208
296 KTPORLGBGLBSWCFCYOSIRPTQMLGSLNIDMASAAFIPLPVIEFVAQLLGKDVLSKP 355
209 ---FOEQVSKELGLVLTXYN---NKTYVDDIDWDQNPKSTFKKADGSEV-SFLEY 260
356 LSDSRVKIKGLRGKVEVTHRANVRKYVAGLTTQPTRELMPFVDENCTMKSIVBYF 415
261 RKQYNQETIDLKQVPL-VSOPKRRRGPGGTIFGPAMLIP-LCYLTGLTDMKRNDFNMVK 318
416 QEMYGFTIQHTHLPCLQVGNCK---ASYLPMERACKIVEGQRYTKRLNEQ----- 463
319 DLAVHTLITPEQOREVGRLLDYI----HKNDNVQRELRDMLGSLFSDSNLFSGRILQTE 374
464 ----ITALLKVTCORAGQRNDILRTVOHNAIDQDPYAKEFGMNISEKLAGSEARILPAP 519
375 --KIHOGGKTFDYNPQFADMSKETRGAPLISVKPLDNWLLIYTRRNYE--AANSLIQNLF 430
520 WLKYHNGEKDKCLPQVGQNMNMNK--KMINGMTVSRWACVNFSPSRVQENVARGFCNELG 577
431 KVTPMQMQRKAIMIEV----DRETE-----AYLRVLQQKVTDADTOIVCLLSNRDK 481
578 QMCVSGMETNPEBVIPIYSGARPQVEKALKHVYHYSNKTGKELELLAILPDNNGSL 637
482 YDAIKKYLCDCTPPSQCWAR----TLGKOQTVMAIATKIALOMNCKMGB-----L 530
638 YGDLKRICEITELGLISOCCLTKHVFKSKQ-----YLADVSLKINVKGGRTVLVDNIS 692
531 WRV-----DIPKLVMVIGDICYHMTA--GRRSIAGFVASIN-EGMTRWFSCIFQDRGQ 583
693 CRPLVSDIP---TIIFGADVHPENGESSPSIAAVVASQDMPVETIKYAGLVCAQAHQ 749
584 ELVDGLKVCVLQAAALR-----ANWSCNEYMPSRITIVYRGVGVGOLKTVNVE 630
750 ELIQDLKTKWQDPVRGTVSGMTRDILLISFRKATGQKPLRIIFTRDGVSGEQFQVLLIYE 809
631 VPQFLDCLKSIGRGYNPRLTVIVYKKEVNTFRFFA-----QSGRLQNPFGTVIDVEVT 684
810 LDAIRKACASLEPNYQPPVFIYVQKEHHHTLFPANNHRDKNSTDNRSGNILLPGTVVDTKIC 869
685 RPWYDFFIYSQAVRSGSVSPTHNVNIVYDNGSLKPDHIQRLTYKLCHIIYNNPGVIRKPA 744
870 HPTEFFYLCSHAGIQGTSRAHVHLWDBNNFTADGIQSLTNLCLTYARCTRSIVP 929
745 PCOYAHKLAPLVGQSIHREPNLISLN 770
930 PAYAYHAAAPRA--RFYLEPEINODN 953

```

RESULT 9
A84668
Argonaute (AGO1)-like protein [imported] - Arabidopsis thaliana
C/Species: Arabidopsis thaliana [mouse-ear cress]
C/Date: 02-Feb-2001 #sequence_revision 02-Feb-2001 #text_change 16-Feb-2001
C/Accession: A84668
R./Lin, X.; Kaul, S.; Rounsley, S.D.; Shea, T.P.; Benito, M.I.; Town, C.D.; Fujii, C.Y.;
M.; Koo, H.; Moffat, K.S.; Cronin, L.A.; Shen, M.; VanAken, S.E.; Umayam, L.; Tallon, L.;
euss, D.; Nierman, W.C.; White, O.; Eisen, J.A.; Salzberg, S.L.; Fraser, C.M.; Venter, J.
Nature 402, 761-768, 1999
A./Title: Sequence and analysis of chromosome 2 of the plant Arabidopsis thaliana.
A./Reference number: A84420; MUID:20083487; PMID:10617197
A./Accession: A84668
A./Status: preliminary
A./Molecule type: DNA
A./Residues: 1-930 <STO>
A./Cross-references: GB:AE002093; NID:G3885334; PIDN:AAC77862.1; GSPDB:GN00139
C./Genetics:
A./Gene: At2g27040
A./Map position: 2
C./Superfamily: rabbit translation initiation factor eIF-2C

Query Match	11.9%	Score 488.5;	DB 2;	Length 930;
Best Local Similarity	21.3%	Pred. No. 3,7e-27;		

Matches	185;	Conservative	170;	Mismatches	340;	Indels	173;	Gaps	34;
Qy	20	GSSGIIIVLSNTHRL-TSRP	WALYOH	-----DYNPLWEARRLS	ALLFO-----	HE	68		
Db	63	GTRQKPTLLNTHFKVDV	ANLQGHFFYS	VALFVDDGRP-VEQ	GVGVRKILDKVH	QTYHS	121		
Qy	69	DLIGKCHAFDG-----	TILFLPKLQ-QKV	TEVFSKTR-----	NG-----	102			
Db	122	DLDGKFAVDGEKILFY	IGALPSNMKDFSV	LVEVSATSKDFV	SRANGSPNGNES	PSD	181		
Qy	103	-----EVRITITL	TNLP-----	PTSPCTQLQV	NIIFRRLKIM	137			
Db	182	GDRKRLRRPNRSKFR	VEISYAAKIP	LOALANAMRGO	SESNQEAIRVLD	ILILRQHAARQ	241		
Qy	138	NLQOIGRNYNENDPIDP-	SHRLVIWPGFT	SILLOYENSIM	LCTDV-SHKVLR	SETVL	194		
Db	242	GCLIVRGSPFH-NDPT	NCBPGVGNILG	CRGFHSFR	TQGGMLNDVMT	TIILKPGPVV	300		
Qy	195	DFMFNFYQTEEHKFQ-	EQVSKELIGL	VLITKYNNTYR	VDDIDWQNKST	TFK-----	247		
Db	301	DFLIANQWARDPSID	WSKAKRTLKRL	RVKVSPSGQ	EFKITGLSDPC	RQTELKKNP	360		
Qy	248	-----KADGSEV	SFLYYRK-----	QYNBEITDLK	OPVLVSQPKRR	PGGTLPGPAMLI	297		
Db	361	NENGFEETTEVT	ADYPRDTRHID	LQYSADLP	PCIN-----	VGKPKR-----	PTIIP	406	
Qy	298	PELC-----	YLTGTLDKMR	NDPNVMKDL	AVHRLTPEQR	QREVGLDIYIHK	NDVQR	350	
Db	407	LECALVPLORYK	ALTFFORS-----	ALVEKSRQ	POERMVTL	SKALKV-SNYDAEP	458		
Qy	351	ELRDWGLSFDSN	LLSPGRILQ	TEKIHQ--GGK	TFDYNPOFAD	WSKETRGAPLIS	VKPLD	408	
Db	459	LLRSCGISISSN	TFQVGRVLP	APKLKMGCS	ETFFRNGE	WNKKE-----	FVEPTKIQ	513	
Qy	409	NWLLIYTRRYEA--	ANSLIQNL	FKVTPAMQ	MOKAIMIE	VDD-----	REAY	455	
Db	514	RWVV-----	NFSARCNVR	QVVDLLIK	GGSGKTEIAS	PFQVFEQ	QFRAPP	MI	569
Qy	456	LRVLQQVKTADT	QIVCLLSNR-K	KYDAIKKYL	CTDCTPSQ	CVVARTLGHQ	QVMAI	514	
Db	570	FKDIQSKLP	QVPOFILCVLP	DKNSDLYG	FWKKKNLTF	GIQVTCMA	PTQPNQYL---	626	
Qy	515	ATKIALQMNKMG-	-ELNRVD-----	IPLKLVMI	VGIDCYHDM	TAGRSIAGFV	562		
Db	627	-TWLLKINAKL	GLSNLSVEST	PAFTVISKP---	TIILGMDVSHG-	SPGQSDVPSTA	681		
Qy	563	ASINEG-----	MTWRFRCITP	DRQBELV	DGL-----	KVCLQALRA	WNSCNEYMS	609	
Db	682	AVVSREWPLIS	KYRASVTRQ	PSKAEMISL	VKNKGTDDG	IIKELLVD	FYTTSSNKRKPE	741	
Qy	610	RIIVYRDGVDG	OLKTLVNV	VEPFLDKLS	IGRGYNPLT	VIIVVKRVN	TRFFAQSGR	669	
Db	742	HLIIFRDG	SVESQFNOVL	NIEDQII	EACLKLDAN	WNPKFLLLV	QAQNHHTKFPQ	TSP-800	
Qy	670	LQNPFGT	IDVETRP	EWDFEYISQ	VRGSVSP	THYNNIV	DNLSGLKPDH	IKRLTYKL	729
Db	801	-ENVP	PGTIIDNKI	CHPKND	DFLCAHAG	MI GTTRP	THVILYDEL	GFSADELQELVHSL	859
Qy	730	CHIYYNWP	GVRVP	APCAV	AKIAFLV	NG	757		
Db	860	SYVQRST	SAISV	APICV	AHLAAALG	887			

RESULT 10
JC6569
translation initiation factor eIF-2C - rabbit
C:Species: Oryctolagus cuniculus (domestic rabbit)
C:Date: 05-Dec-1998 #sequence_revision 05-Dec-1998 #text_change 21-Jul-2000
C:Accession: PC6505; JC6569
R:Zou, C.; Zhang, Z.; Wu, S.; Osterman, J.C.
Gene 211, 187-194, 1998
A:Title: Molecular cloning and characterization of a rabbit eIF2C protein.
A:Reference number: JC6569; MUID:96267198; PMID:9602122

265	Db	DLEGGK-EMWTGPFSSAHIASNYRPLLNIDVAHTAFYKTRITVLFQFMCVDVLERTSKPNR	323
204	Qy	-----	TEEH 207
324	Db	NNPRGPGAGYRGGRGGARGGSYQNFNGRPPGANVRDDFGGNGLTFTMDTLSDTQLS	383
208	Qy	KFOBOYSKELL-GLVVLTKYNNKTYRDDDDWDQNP--KSTFKKAD--GSEV--SFLFVY	260
384	Db	SFTRIFGDSIRGMKIRATHRPNRAIRYKYNLSQLPADKLMFOGIDEEGRQVVCVADVF	443
261	Qy	RKQVNOEITDLKOPVLVSQPKRRRPGGTLPGPAMLIP-BEC-----YLTGLTKDMN	312
444	Db	SEKVG----PLKYPKL---PCLHVGP---PTRNIFIMEHCLIDSPOKYNKMKTEKOTS	492
313	Qy	DFNVMKDLAVHTRLTPQRQREVGRLLIDYTHKNDNVQRELEDDWGLSFDNSLLISGRIILQ	372
493	Db	--AIIKAAAVDA---TQREDRIKOLA--AQASFGTDPFLKEFGVAYSSQMIETSAIRVIQ	544
373	Qy	TEKHOGGKTFDYN----POFADWSKETRGAPLLISVKPDLNWLILYTRRNYEAAANSLION	428
545	Db	PPPIPMFEGGNRRSINPVVFPFDGGSWMDHQTLYXPATCRYSMIALVDPDRDQTSLOTFCQS	604
429	Qy	LEXYTBAACWMDOK-----NMTVEYDQDPMAYRDUKQCKTMDCT	470

Db

```

473 QY 473 LLSSNEKKDYDAIKKYLCTDCTPSCQVARTLGKQGTWMAIKIALOMCKYGG----- 528
656 Db 656 VLOKNSDIYNTVKQESDIVHGIMSQCVLMKNVSR--PTPATCANIIILKLANMKVGGINSR 713
529 QY 529 -----ELWRVDIPLKLVMIIVGIDCVHDMTAGRR-----STAGFVASINEGMTRWFSRCI 577
714 Db 714 IVADQITNKVLDQP---TWVGIDVTHPTCAERKMMPVAAIIVANVDLLPQSYGANVK 770
578 QY 578 FQDRGOELVDGLKVCICQALRAWNSCNEYMPSRIIVVRDGVGDGQLKTLVNVYVPPQFLDC 637
771 Db 771 VQKKCESVYLLDAIRERITITRYTKQKPARIIIVVRDGVSGQCFSEVLREEIQSINTA 830
638 QY 638 LKISIGRGYNPRLTVIVVKKRVNTRFFAQ-----SGGRLOQLPGTVIDVETVRPEWYDFPI 693
831 Db 831 CLAIABDFRPPIYTVVQKHHARIICKFPNDMVGRKAKNVPFGTTVTGTGIVSPGPGDFYL 890
694 QY 694 VSOAVRSGSVSPHYNNVVDNSGLKPDHIOQLRYKLCIIYYNNMPGIVRVPAPOYAHKLA 753
891 Db 891 CSHYVGQGTSRPARYHVLLDECKFTADEIQNIITVGMCHTYGCRTRSVSIPTPVYYADIVA 950
754 QY 754 FLVGQSIIHREPNIISLN 770
951 Db 951 TRARCHIKRKLGLADNN 967

RESULT 13
T23165
hypothetical protein T23B3.2b - Caenorhabditis elegans
C:Species: Caenorhabditis elegans
C:Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 04-Mar-2000
C:Accession: T23165; T25100
R:Cottage, A.
submitted to the EMBL Data Library, January 1996
A:Reference number: Z19701
A:Accession: T23165
A:Status: preliminary; translated from GB/EMBL/DDBJ
A:Molecule type: DNA
A:Residues: 1-1035 <WIL>
A:Cross-references: EMBL:Z68750; FIDN:CAA92970.1; GSPDB:GN00022; CESP:T22B3.2b
A:Experimental source: clone K01A6
R:Lennard, N.
submitted to the EMBL Data Library, December 1995
A:Reference number: Z19981
A:Accession: T25100
A:Status: preliminary; translated from GB/EMBL/DDBJ
A:Molecule type: DNA
A:Residues: 1-1035 <WIL>

```

C;Species: Caenorhabditis elegans
C;Date: 06-Jan-1995 #sequence_revision 06-Jan-1995 #text_change 18-Feb-2000
C;Accession: S41013

R;Thomas, K.
submitted to the EMBL Data Library, December 1993
A;Reference number: S41011
A;Accession: S41013
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-958 <THO>
A;Cross-references: EMBL:729121
C;Genetics:
A;Introns: 52/2; 350/2; 422/3; 592/1; 679/3; 721/2; 887/2; 929/1
C;Superfamily: rabbit translation initiation factor eIF-2C

Query Match 10.3%; Score 421; DB 2; Length 958;
Best Local Similarity 21.6%; Pred.No. 3.2e-22;
Matches 199; Conservative 148; Mismatches 367; Indels 206; Gaps 369

Qy	20	GSGGIIVRLSTNHFL--TSRPQWALYQHID--YNPLMEARLL---RSALLP-----OHED	69
Dd	11	GTIGRQIPVKSFFAMDLKNPMGVVIQHVBIHP--GCRKLDKDEMIIFWKAVSDHNP	68
Qy	70	LICK--CHAFDGT--ILFLPKILO-----OKTEVFSTRNGEDVRIT---	109
Dd	69	IHNKFAIADAGHQLTYVAELPDGGSVRLDCEASLPKNDRTCAISIQNVGFVL	128
Qy	110	----TLNLEPPPTSCFLOFYNIIFRLLIKIMLQIGRNYYN-----PNMPI	153
Dd	129	LEMQRTRINLDERVLTPIQILDIIICQSGLCPILKN-SANFYTWKSSCYRIPTAQAL	187
Qy	154	DFPSRLVINGFTSIILOYEN-SIMCLTDVSHKVLASE--TVLDFMNFYHQ-----	203
Dd	188	DLEGKK-EMWTGFTSSAHIASNYRPLLIDIVAHTAFYKTRITVLQFMCDVINERTSKPNR	246
Qy	204	-----T 204	
Dd	247	NNPRGPGGPGGVGRGGRGGCGYSTGNFRGPPGANVRDDPGGNGLFTMTDLSRDT	306
Qy	205	EBHKFQBOYSKELI--GLVYLTKYNNKTRYDVDDIDWDQNPF-KSTFKAAD--GSEV--SFL	257
Dd	307	QLSSPETRIFGDIAIRGMKIRAHPNAIRVVKVNSLQLPADKLMFQGIDEEGRVVCSPA	366
Qy	258	EYRKQYNQEITDLKPVLINSQPERRRGPGGTLPGPAMLIP-ELC-----VLTGLTDK	309
Dd	367	DFSEKYG-----PLUKYPKL-----PCLHWGP-----PTRNIFLPMHECLIDSQKYNKMSKX	415
Qy	310	MENDFNVMKOLAVHRTLTPGEORQEVGRLDIYIHKNDNVQRELRDWGLSPDSNLLSFSGR	369
Dd	416	QTS--ALIKAADVDA---TOREDRIKOLA--QAASGTDPFPLKEFGVAUSSOMIOITTA	467
Qy	370	LIOTBKIHQQGKTFDYN-----POFADMWSKETRGAPLISVKPLDNWLLIYTRRNYEANS	425
Dd	468	VIQPPPIFMFGNNNSVPVFPFKDGSGTMDNQTLMPATCRSYSMIALVDPRDOTSLQTF	527
Qy	456	IQNLKFVTPMGMMQMRK-----AIMLEVDDRTEAYLRVLQOKVTADT-QI	469
Dd	528	CSQLTNKATAMGNFPWPDLIVKGRSKBDVCTLFTEIAD-----EYRVNTVDCD	578
Qy	470	VVCLSSNRKDYDAIKKLYLCDTCPSPQCQVARTLGKQQTVMATKIALONCKRWGG-	528
Dd	579	IIVLVQSNKSDIYTWTVKEQSDIVHGIMSQCVLMKVSR--PIPATCANIVLKLNKXGGI	636
Qy	529	-----ELMRVDIPLKLMIVIGDICVHDMTAGR-----SIAGFVASINEGMTWNFS	574
Dd	637	NSRIADVKITNKYLVDOF-----TMVVGDIVTHPTQAEWRNMNPSVAAIVANDLLPOSYGA	693

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: December 4, 2003, 19:08:48 ; Search time 18 Seconds

(without alignments)
2024.761 Million cell updates/sec

Title: US-10-043-774B-2

Perfect score: 4102
Sequence: 1 MIFGNTFRQLDHVKESKGTG.....VQGSIHREPNSLSNRLYYL 775

Scoring table:

Gapop 10.0 , Gapext 0.5

Searched: 127863 seqs, 47026705 residues

Total number of hits satisfying chosen parameters: 127863

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : SwissProt_41.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	DB ID	Description
1	557.5	13.6	834	1	IF2C SCHPO
2	510	12.4	1048	1	AG01 ARATH
3	506.5	12.3	857	1	I2C1 HUMAN
4	504.5	12.3	863	1	I2C2 RAT
5	503	12.3	988	1	P1NE ARATH
6	495.5	12.1	987	1	AG0L ARATH
7	482	11.8	813	1	I2C2 RABIT
8	421	10.3	1040	1	YO43 CAEEL
9	374	9.1	377	1	I2C2 HUMAN
10	248.5	6.1	1032	1	YO53 CAEEL
11	137.5	3.4	1912	1	DICE HUMAN
12	134.5	3.3	1906	1	DICE MOUSE
13	118	2.9	1120	1	MFD RICPR
14	114	2.8	2035	1	HMP2 YEREN
15	113.5	2.8	560	1	PMAC PSOL
16	113	2.8	949	1	PMAC ARATH
17	113	2.8	3144	1	VP13 YEAST
18	108	2.6	1317	1	N145 YEAST
19	108	2.6	1653	1	CLH YEAST
20	106.5	2.6	690	1	Y173 URBPA
21	104	2.5	1180	1	C4AA BACTI
22	104	2.5	1277	1	PD55 YEAST
23	104	2.5	4128	1	PRD5 YEAST
24	103	2.5	1343	1	RPOB HAIN
25	103	2.5	1575	1	IQG2 HUMAN
26	102.5	2.5	354	1	CHLI CHLVU
27	102.5	2.5	869	1	YD95 YEAST
28	102.5	2.5	1170	1	ITR2 BOVIN
29	102	2.5	3122	1	DPO2 MOUSE
30	101.5	2.5	3135	1	S230 PLAF0
31	101	2.5	1388	1	RPOD TOBAC
32	101	2.5	1886	1	POL COYMY
33	100.5	2.5	1162	1	VGL2_IBVE

34	100.5	2.5	1169	1	EX5B BORBU
35	100	2.4	963	1	KINH HUMAN
36	100	2.4	1014	1	NEBL HUMAN
37	100	2.4	1587	1	TOP2_PENCH
38	99.5	2.4	879	1	MISP_DROME
39	99.5	2.4	1358	1	SIR4 YEAST
40	99.5	2.4	4568	1	DYHC CAEEL
41	99	2.4	749	1	STA4 MOUSE
42	99	2.4	806	1	SVL_BACHD
43	99	2.4	944	1	DNL4 YEAST
44	99	2.4	1086	1	SVI RICPR
45	99	2.4	2670	1	IP3T RAT

ALIGNMENTS

RESULT 1

IF2C SCHPO

ID IF2C SCHPO STANDARD; PRT; 834 AA.

AC 074957;

DT 16-OCT-2001 (Rel. 40, Created)

DT 16-OCT-2001 (Rel. 40, Last sequence update)

DT 28-FEB-2003 (Rel. 41, Last annotation update)

DE Probable eukaryotic translation initiation factor 2C (eIF2C) (eIF-2C).

GN SPCC736.11.

OS Schizosaccharomyces pombe (Fission yeast).

OC Eukaryota; Fungi; Ascomycota; Schizosaccharomycetes;

OC Schizosaccharomycetales; Schizosaccharomycetaceae;

OC Schizosaccharomycetes.

OX NCBI_TaxID=4896;

RN [1]

RP SEQUENCE FROM N.A.

RC STRAIN=972;

RX MEDLINE=21848401; PubMed=11859360;

RA Wood V., Gilliam R., Rajandream M.A., Lyne M., Lyne R., Stewart A., Sgouros J., Peat N., Hayles J., Baker S., Basham D., Bowman S., Brooks K., Brown D., Brown S., Chillingworth T., Churcher C.M., Collins M., Connor R., Cronin A., Davis P., Feltwell T., Fraser A., Gentles S., Goble A., Hamlin N., Harris D., Hidalgo J., Hodgson G., Holroyd S., Hornsby T., Howarth S., Huckle E.J., Hunt S., Jagels K., James K., Jones L., Jones M., Leather S., McDonald S., McLean J., Mooney P., Mungle S., Mungall K., Murphy L., Niblett D., Odeh C., Oliver K., O'Neill S., Pearson D., Quail M.A., Rabinowitsch E., Rutherford K., Rutter S., Saunders D., Seeger K., Sharp S., Skelton J., Simmonds M., Squares R., Squares S., Stevens K., Taylor K., Taylor R.G., Tivey A., Walsh S.V., Warren T., Whitehead S., Woodward J., Voiclaert G., Aert R., Robben J., Grymonprez B., Welljans I., Vanstreels E., Rieger M., Schaefer M., Mueller-Auer S., Gabel C., Fuchs M., Fritz C., Holzner E., Moestl D., Hilbert H., Borzym K., Langer I., Beck A., Leirach H., Reinhardt R., Pohl T.M., Eger P., Zimmermann W., Wedler H., Wambutt R., Purnelle B., Goffeau A., Cadieu E., Dreano S., Gloux S., Lelaure V., Mottier S., Galibert F., Aves S.J., Xiang Z., Hunt C., Moore K., Hurst S.M., Lucas M., Rochet M., Gallard C., Tallada V.A., Garzon A., Thode G., Daga R.R., Cruzado L., Jimenez J., Sanchez M., del Rey F., Benito J., Dominguez A., Revuelta J.L., Moreno S., Armstrong J., Forsburg S.L., Cerutti L., Lowe T., McCombie W.R., Paulsen I., Potashkin J., Shipakovski G.V., Ussery D., Barrell B.G., Nurse P.

RT "The genome sequence of Schizosaccharomyces pombe."

RL Nature 415:871-880(2002).

CC -!- FUNCTION: Plays an important role in the eukaryotic peptide chain initiation process (By similarity).

CC -!- SIMILARITY: BELONGS TO THE ARGONAUTE FAMILY.

CC -!- SIMILARITY: Contains 1 PAZ domain.

CC -!- SIMILARITY: Contains 1 Piwi domain.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/>)

CC or send an email to license@isb-sib.ch.

```
CC -----
CC EMBL; AL023705; CAA19275.1; -.
CC PIR; T41568; T41568.
CC GenDB SPombe; SPC0736.11; -.
CC InterPro; IPR003100; PAZ.
CC InterPro; IPR003165; Pw1.
CC Pfam; PF02170; PAZ; 1.
CC Pfam; PF02171; Pw1; 1.
CC PROSITE; PS00821; PAZ; 1.
CC PROSITE; PS00822; Pw1; 1.
KW Initiation factor; Protein biosynthesis.
FT DOMAIN 212 326 PAZ.
FT DOMAIN 500 799 Pw1.
SQ SEQUENCE 834 AA; 94438 MW; F5CE397EF590D9E4 CRC64;

Query Match          13.6%; Score 557.5; DB 1; Length 834;
Best Local Similarity 22.9%; Pred. No. 1.2e-32;
Matches 198; Conservative 155; Mismatches 346; Indels 167; Gaps 31;

QY 20 GSGIIVRLSTNHRFRITSPQWALYQYHIDYNFLMEARRLSALLFOHEDL----- 70
Dd 16 GGLGKQITLKANFFQIISLPNETINQYHIVGDSRVPKQSLIWNKSEVKYQFGSSWM 75
QY 71 -----IGKCHAFDGTILF-----LPKELQ---QKVTEV-----PSKTRNGED 104
Dd 76 NSVYDGRSMCWSGDADGTIKVNIAGESHPREIFSIQKSKINLHTLSQFVNSKYSDD 135
QY 105 VRITITLNLPTSPSTCLQFYNIIFRR-----LLKIMNLQIQGRVYNPNPDIDIPSHR 159
Dd 136 -----PQVLSIMFLDLILKKPSETLFGFMHSFFTGENGVSLLGGVVE----- 178
QY 160 LVTPGFTSILQYENSIMLCTDVSHKVL-RSTVLDFM----- 197
Dd 179 --AWKGFQYSIRENCGFMSVNVDISSAFWNDSLLQILMEYDCSNVRDLTFDLKRLS 236
QY 198 --PNFVHQTEEHKFOQVSKELGLVLTQYNNKTYRVDIDWDQPKSTF-KKADGSE- 253
Dd 237 RKRFLEKLVTCQHRNN-----VGTDLANRVYSIEGFSKSDSDFVRLNGEEQ 285
QY 254 -VSFLEYRYKQYNOEITDLKQPLVSPQRRRRPGGTLPGPAMLIPELCVLT---GLTK 309
Dd 286 KISVARYFLENHVRVLYQYNLPCL-----VKGAMLPTEFVVKQRYTK 333
QY 310 MRNDFNMKDLAVHTRLTPEQREYGRGLDYTHKND-NVQRELRLDWGLSFDNLSFSG 368
Dd 334 LNSD-----QTANMIRFAVQRPPEPVQIQIDDFVHQMDWDTPVLTQYGMKIQKMLEVEA 388
QY 369 RILQTEKIHQGGKTFDYNQPADMSKETRCAPLISV--KPLDNWLLI---YTER-NYEA 422
Dd 389 RVLETPSIRYGGCIE-RPVSGRWN--LRGRFLDPPRAPIRSNVAVWCFTSTRRLPMRGI 445
QY 423 NSLIQNLKFKVTPAMGMQ--MRKAIMIEVDNR--TEAYLRVLOQV---TADTQIVVCLL 474
Dd 446 ENFLQYVQVLTSLGFINFMVKPPLVYADIRGSVEELCTLYKKAQVGNAPDYLFFIL 505
QY 475 SSNRKDYDAIKYLCTDCTPTSCQVAVPLGKQQTMAITKIALQMNCKMG---ELW 531
Dd 506 DKNSPEFYSGIKVCNMTLGVSPQCAISKHI--LQSKPQYCANLGMKINVKVGGINCSLI 563
QY 532 RVDIPILKLV--MIVGIDCVH-DMTAGRRSTAGFVASINEGWTWFSRCIFQDRGQELVDG 588
Dd 564 PKNPLGNVPTLIGGVVHPGATGVSTASIVASVDLNGCKYTAVSRSQPRHQVEIEG 623
QY 589 LKVCLOALRAWNSCNEYMPSRIIVRVGDVGDLTKLVNVEYVQFQDLCLKSTGRGNPR 648
Dd 624 MKDIVVILQGFRAWKQQRRIIYFRDGTSEGOFLSVINDELISOKEACHSLSPKYNK 683
QY 649 LTVIVVKRVNTRPFAQS-----GRLGNPLPGTVIDVEVTPPEWDFPFIQSVAVRSGVSP 705
Dd 684 ILVCTTQKRHRHFFIKNKGSDGRGNPLPGTIIKHTVHPYQYDFLISHPSLQGVSP 743
QY 706 THYNYVDNSGLKPDHTRITQYKLCIYYNPNVGVIRVPAPQVAKHLAFL----- 755
```

```
Db 744 VHYTVLHDEIQMPDQFQTLNCLYVYATRSVSLVPEVYIAHLVSNLARYQDVTADD 803
QY 756 -----VGQSIHREPNLSLNL 772
Dd 804 TFEVTESEASMDQEV--KPLALSLSKL 827

RESULT 2
AG01 ARATH
ID AG01 ARATH STANDARD; PRT; 1048 AA.
AC O04379.
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Argonaute protein.
GN AG01 OR AT1G48410 OR FLA17.3.
OS Arabidopsis thaliana (Mouse-ear cress).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
OC eurosids II; Brassicales; Brassicaceae; Arabidopsi.
NCBI TaxID=3702;
RN [1]
SEQUENCE FROM N.A.
RC STRAIN=cv. Columbia; TISSUE=Leaf;
RX MEDLINE=98090460; PubMed=9427751;
RA Bohmert K., Camus I., Bellini C., Bouchez D., Caboche M., Benning C.;
RT "AG01 defines a novel locus of Arabidopsis controlling leaf
development.";
RL EMBO J. 17:170-180(1998).
RN [2]
SEQUENCE FROM N.A.
RC STRAIN=cv. Columbia;
RX MEDLINE=21016719; PubMed=11130712;
RA Theologis A., Ecker J.R., Palm C.J., Federspiel N.A., Kaul S.,
RA White O., Alonso J., Altafi H., Araujo R., Bowman C.L., Brooks S.Y.,
RA Buehler E., Chan A., Chao Q., Chen H., Cheuk R.F., Chin C.W.,
RA Chung M.K., Conn L., Conway A.B., Conway A.R., Creasy T.H., Dewar K.,
RA Dunn P., Egu P., Feldblum T.V., Feng J.-D., Fong B., Fujii C.Y.,
RA Gill J.E., Goldsmith A.D., Haas B., Hansen N.F., Hughes B., Huizar L.,
RA Hunter J.L., Jenkins J., Johnson-Hopson C., Khan S., Khaykin E.,
RA Kim C.J., Koo H.L., Krenetskaia I., Kurtz D.B., Kwan A., Lam B.,
RA Langin-Hooper S., Lee A., Lee J.M., Lenz C.A., Li J.H., Li Y.-P.,
RA Lin X., Liu S.X., Liu Z.A., Luros J.S., Maiti R., Marziani A.,
RA Miltschker J., Miranda M., Nguyen M., Nierman W.C., Osborne B.I.,
RA Pai G., Peterson J., Pham P.K., Rizzo M., Rooney T., Rowley D.,
RA Sakano H., Salzberg S.L., Schwartz J.R., Shinn P., Southwick A.M.,
RA Sun H., Tallon L.J., Tambunga G., Toriumi M.J., Town C.D.,
RA Utterback T., Van Aken S., Vaysberg M., Vysotskaia V.S., Walker M.,
RA Wu D., Yu G., Fraser C.M., Venter J.C., Davis R.W.;
RT "Sequence and analysis of chromosome 1 of the plant Arabidopsis
thaliana.";
CC 1- FUNCTION: ESSENTIAL FOR PROPER DEVELOPMENT OF LEAVES AND FLORAL
CC ORGANS AND FORMATION OF AXILLARY MERISTEMS.
CC 1- SUBCELLULAR LOCATION: Cytoplasmic (Potential).
CC 1- TISSUE SPECIFICITY: WIDELY EXPRESSED AT LOW LEVELS.
CC 1- DEVELOPMENTAL STAGE: EXPRESSED THROUGHOUT ALL DEVELOPMENTAL
CC STAGES.
CC 1- SIMILARITY: BELONGS TO THE ARGONAUTE FAMILY.
CC 1- SIMILARITY: Contains 1 PAZ domain.
CC 1- SIMILARITY: Contains 1 Pw1 domain.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC EMBL; U91995; AAC18440.1; -.
CC EMBL; AC007932; AAC49755.1; -.
DR
```

```
DR InterPro; IPR003100; PAZ.
DR InterPro; IPR003165; Piwi.
DR Pfam; PF02170; PAZ; 1.
DR Pfam; PF02171; Piwi; 1.
DR PROSITE; PS0821; PAZ; 1.
DR PROSITE; PS0822; PIWI; 1.
KW Developmental protein.
FT DOMAIN 391 501 PAZ.
FT DOMAIN 676 997 PIWI.
FT DOMAIN 13 104 GLY-RICH.
SQ SEQUENCE 1048 AA; 116190 MW; 385146343A09C541 CRC64;

Query Match 12.4%; Score 510; DB 1; Length 1048;
Best Local Similarity 25.1%; Pred. No. 4.8e-29;
Matches 213; Conservative 136; Mismatches 376; Indels 124; Gaps 33;

QY 20 GSSGIIVRLSTNHFRSLTSPQWALYQYHIDYNPLMEARLRSLALPQHEDLIGKCH----75
Db 186 GQSGKRCIVKANHF-FAELPKDLHHYDVTITEVTSRGVNRVAVMKQLVDNYRDSHLSGR 244
QY 76 --AFDG-----TILFLP-----KRLQOKVTEVSKTRNGEDVRIIT 109
Db 245 LPAYDGKSLYTAGLPFNFSKEPRINLLDEEVGAGQRREREPKVIKLVARADLHLGM 304
QY 110 TLNLEPPTSPCTCLQFYNIIFRLLKIMNLQOIGRNYNPNDPIDIPSHR-----LVWIP 164
Db 305 FLEGKSDAPQALQVLDIVLR-LTSRYIPVGRSFYSP-----DICKQSLGDGLESWR 359
QY 165 GFTTSILOYENSIMLCTDVSHKV-LRSETVLDPMFNFYHOTEHK-----FQGVSKELI 218
Db 360 GFYOSIRPTQMLSLNIDMSSTAFIEANPVIQVCDLLNDRISRPDSADRVKIKKALR 419
QY 219 GL-VLITKYN--KTRVDDIDWDQPKSTFKAD-GSEVSFLEYKQYNQBITLQKP 274
Db 420 GVXVEVTHRGNNRKYRISGLTAVATRELTPPVDERTQKSVVEVPHETVGFRIQHTQLP 479
QY 275 VLVSQPKRRRGPGCTLPGPAMLPCLCYLT--GLTDQRNDNMVMDLAVHTRLTPEQRQ 332
Db 480 CLOVGNSNR-----PNYLPMEVCKIVGQRYSKNLNERQITALLKV-TQRPIDRE 529
QY 333 REV---GRLLDYTHKNDNVORELDRMGLSPDSNLLSPSRILOTE--KIHQGGKTFDYNP 387
Db 530 KDILQTVQLNDY--AKDNYAQE---FGIKISTSLASVEARILPPPMKLVHESGREGTCLP 584
QY 388 OFADWSKETRGAPLISVKPLDNLMLLYITRNYE--AANSLIONLFKVTTPANGQMR-KAI 444
Db 585 QVGQWNNMKN--XKINGGTNNMICINFSRQVQDNLARTFCBELAQCYVSGHAFNPEPV 642
QY 445 MIEVDRTTEAYLVLOQKVTAQT-----QIVVCLLSNRKKDYDAIKKYLCTDCPT 495
Db 643 LPVVSARPEQVEKVLKTRVHDATSKLSQGEIDLLIIVLPDNGSLYGLDKRICETELGI 702
QY 496 PSQCQVARTLGK-QQTWAIATKIALQMKMG-ELMRVD-----IPL---KLVMIVGI 545
Db 703 VSQCCLTKHFKMSQYMA---NVALKINVKVGRNTVLVDALSRRIPFLVSDRPTIFGA 759
QY 546 DCVHDMTA--GRSISAGFVASIN-BGTRWFSRCIFQDRGOELVDGI-----589
Db 760 DVTHPHGEDSSPSIAAVASQDWPETTKYAGLVCAQHQCQLIQDLFKEWKDPQKGVVT 819
QY 590 KVCQLQALRAWNSCNEYMPRIIVYRGVGDGQLKTLVNTVEVQFLDCLKSIKRGYNPLR 649
Db 820 GMIKELLIAFRSTGCHKPLRIIFYRGVSEGQYQVLLVELDAIRKACASLENGYQPV 879
QY 650 TVIIVKRVNTRFFAQ-----SGGRQLNPLPGTVIDVETRPEDWYDFFITVQAVRSGSV 703
Db 880 TFFVVKRHHRLFAQNHNDHRSVDRSGNILPGTVVDSKICHPTEPDFYLCSHAGIQGTS 939
QY 704 SPTHYVINYNSGLKPDHIOQLTKLCHIYNNWPGVTRVAPCOYAKHLAFLVQSQSHRE 763
Db 940 RPAHYHVLWDENNFTADGLQSLTNLCTVARTCRSVISVPPAYYAHAAFA--RFYME 997
QY 764 PNLSLSNRL 772
```

```
Db 998 PETSDSGSM 1006
RESULT 3
ID I2C1 HUMAN STANDARD; PRT; 857 AA.
AC Q9ULJ8;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Eukaryotic translation initiation factor 2C 1 (eIF2C 1) (eIF-2C 1)
DE (Putative RNA-binding protein Q99).
GN EIF2C1
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=20005943; PubMed=10534406;
RA Roosters R., Adams V., Betts D., Moos R., Schmid M., Siermann A.,
RA Haastem S., Weitz S., Lichter P., Heitz P.U., von Knebel Doeberitz M.,
RA Briner J.;
RT "Human eukaryotic initiation factor EIF2C1 gene: cDNA sequence,
RT genomic organization, localization to chromosomal bands 1p34-p35, and
RT expression.";
RL Genomics 61:210-218(1999).
CC -!- FUNCTION: Plays an important role in the eukaryotic peptide chain
CC -!- INITIATION PROCESS.
CC -!- SIMILARITY: BELONGS TO THE ARGONAUTE FAMILY.
CC -!- SIMILARITY: Contains 1 PAZ domain.
CC -!- SIMILARITY: Contains 1 Piwi domain.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; AF093097; AAF00068.1; -.
CC Genew; HGNC:3262; EIF2C1.
CC MIM; 606228; -.
CC GO; GO:0003850; C:eukaryotic translation initiation factor 2...; TAS.
CC GO; GO:0003743; F:translation initiation factor activity; TAS.
CC GO; GO:0006446; P:regulation of translational initiation; TAS.
CC InterPro; IPR003100; PAZ.
CC InterPro; IPR003165; Piwi.
CC Pfam; PF02170; PAZ; 1.
CC Pfam; PF02171; Piwi; 1.
CC PROSITE; PS0821; PAZ; 1.
CC PROSITE; PS0822; PIWI; 1.
KW Initiation factor; Protein biosynthesis.
FT DOMAIN 226 346 PAZ.
FT DOMAIN 515 816 PIWI.
SQ SEQUENCE 857 AA; 97214 MW; 1DBE524AE7CBAP66 CRC64;

Query Match 12.3%; Score 506.5; DB 1; Length 857;
Best Local Similarity 24.1%; Pred. No. 6.5e-29;
Matches 203; Conservative 160; Mismatches 347; Indels 133; Gaps 37;

QY 20 GSSGIIVRLSTNHFRSLTSPQWALYQYHIDYNPLMEARLRSLALPQHEDLIG-K 73
Db 30 GTVGKPIKLLANYFEV-DIPKIDVHYEVDIKPKCPRVNREVVEYVQHFKEPFGDR 88
QY 74 CHAFDG-----TILFLP---KRLQOKVTEVSKTRNGED--VRIIT-----T 110
Db 89 KPVIDGKKNYITVATLPIGNERVDFEVT---IPGEGKDRIFKVSIKWLAIYSWRMLHEA 144
QY 111 LTNLEPPTSPCTCLQFYNIIFRLLKIMNLQOIGRNYNPNDPIDIP-SHRLVTPGFTTS 169
```

```

Db 145 LVSGQIPVLESQALVDVAMRHUAS-MRYTPVGRSFFSPGEGYHPLDGGREVFWFHOS 203
Qy 170 ILQYENSIMCLTDSVSKVL-RSETVLEDFMF-----NFVHQTEBEHFORQV--SKELIGL 220
Db 204 VEPAMKMLNIDVSATAFYKAQPVIEFCEVLDIRIDEQPKPLTDSQVRFTKEIGL 263
Qy 221 VVLTKY---NNKTYRVDIDMDGNPKSTF---KKADGSEVSFLEYRKYQNOEITDLKQ 273
Db 264 KVEVTHCGQMKRYRVCNTRRPASHQTPPLQLESQGTVECTVAQYFKQKYNLQ---LKY 320
Qy 274 PVL---VSQPKRRRGGTLPAPMLIPELCVLTG-----LTDKMRDNFNMKDLAV 322
Db 321 PHLPCLOVQEQKH-----TYLPLEVCNIVAGQCIKLTLD---NOTSTM--IKA 365
Qy 323 HTRLTPQEQREVRGLIDYIHKNDNVQRELDRWGLSFDGNLLSFGRIILQTEKHGGKT 382
Db 366 TARSAPD-RQBEISRLMK--NAYNLDVPYIQEFGIKVKDDMTVTGRLPAPILQYGGN 422
Qy 383 FQY-NPQFADWSKETGAPLISVKPLDNMLLIYTRNRYEANSLLION----LFKVTTPAMG 437
Db 423 RAIATENQGVW--DMRGKQFYNGIEIKWAIACFAPQKQCRBEVLKNFTDQLRKISKDAG 480
Qy 438 MOWRK-----ALMEVDDRTYAYLRVLQOKVTADTQIVVCLSSNRKDKYDAIKKYLCTDC 493
Db 481 MFIQGGQCFCKYAQAGSDSVEPMRHLKNTYSG-LQLIIVIL-PGKTPVTAEVKRVGDILL 538
Qy 494 PTPSQCVVARTLGKQOTVMAIATKIALQMNCKMGBELVRDIP-----LKLWMIVGID 546
Db 539 GWATQCVQKNNVK--TSPQTLNGLKINVKLGG-INNLVPHQRSAYFPQPVIFLAD 595
Qy 547 CYHDMTAG---RRSIAGFVASINEGTRMFRSCIFQDRGOELVDGLKVCLOALRAWNSC 603
Db 596 VTHP-PAGDKKPSITAVVSGMDAHPSCYCATVRVORPROEITIEDLSYMRRELLIOFYKS 654
Qy 604 NEYMPSSIIYVRDGVGGQGLKTVNVEVPOFLDCLXSIGHGYNPLTVIVKRVNTRFE 663
Db 655 TRFKPRIIFYRQVVEGQPLQHLVELLAIRDACIKLEKQYQPGIYIVVQGRHRLRF 714
Qy 664 A-----QSGRLQNLPGTVIDVEVTRPEWYDFIVSQAVRSQVSPSTHYNVIYDNSGLXP 719
Db 715 CADKNERIGKSGNIPAGTVDVNIWTHPEFDFYLCSHAGIQGTSRPSHYVVLWDDNRFTA 774
Qy 720 DHIQRLTYKLCHLYNNWGVIRVPACQYAHKAF-----LVGOSI 761
Db 775 DELQILTYQLCHTYVRCRSVSIAPAYYARLVAFRARYHLVDKEDSGEGSHISQSGS 834
Qy 762 REP 764
Db 835 RDP 837

```

RESULT 4

```

ID 12C2 RAT STANDARD; PRT; 863 AA.
AC Q9QZ81;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Eukaryotic translation initiation factor 2C 2 (eIF2C 2) (eIF-2C 2)
DE (Goldi ER protein 95 kDa) (GERp95).
DE EIF2C2.
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX NCBI_TaxID=10116;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Sprague-Dawley; TISSUE=Hepatoma;
RX MEDLINE=99443791; PubMed=10512872;
RA Cikaluk D.E., Tabbaz N., Hendricks L.C., DiMattia G.E., Hansen D.,
RA Pilgrim D., Hobman T.C.;
RT "GERp95, a membrane-associated protein that belongs to a family of
RT proteins involved in stem cell differentiation."

```

```

RL Mol. Biol. Cell 10:3357-3372(1999).
CC -!- FUNCTION: Plays an important role in the eukaryotic peptide chain
CC initiation process.
CC -!- SIMILARITY: BELONGS TO THE ARGONAUTE FAMILY.
CC -!- SIMILARITY: Contains 1 PAZ domain.
CC -!- SIMILARITY: Contains 1 Piwi domain.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; AF195534; AAF12800.1; -.
CC InterPro; IPR003100; PAZ.
CC InterPro; IPR003165; Piwi.
CC Pfam; PF02170; PAZ; 1.
CC Pfam; PF02171; Piwi; 1.
CC PROSITE; PS0821; PAZ; 1.
CC PROSITE; PS0822; PIWI; 1.
CC KW Initiation factor; Protein biosynthesis.
CC FT DOMAIN 239 352
CC FT PAZ.
CC FT PIWI.
CC SQ SEQUENCE 863 AA; 97659 MW; 9A1E0D6B84F7B622 CRC64;
Query Match 12.3%; Score 504.5; DB 1; Length 863;
Best Local Similarity 24.1%; Pred. No. 9.2e-29;
Matches 197; Conservative 151; Mismatches 349; Indels 121; Gaps 33;
Qy 20 GSSGIIVRLSTNHPRLTSRPOWALYQYHIDYNPLMEARLSAL---LPQH--EDLIG-K 73
Db 36 GTTGRTIKLQANFFEM-DIPKIDIVHYELDIKPEKCPRVNREIVEHVMVQHEKTFQIGDR 94
Qy 74 CHAFQG-----TILFLPKELQKVTEVFSKTEGND--VRITI-----TLTN 113
Db 95 KPVFGKRNLYTAMPLPIG-RDKLEVLTLPGEGKDRIFKVKISKWVSCVSLQALHDALSQ 153
Qy 114 ELPPTSPTCLQFNNIIFRLLKIMLQOIGRNNPNPIDIP-SHRLVWPGFTTSIQ 172
Db 154 RLPSVFFETIQALDVVMRH-LFSMRYTPVGRSFFTAEGSCSNPLGGREVFHQSVP 212
Qy 173 YENSIMCLTDSVSKVL-RSETVLDP---MNFVHYOTEHKK-----FQEQVSKELIGLVVL 223
Db 213 SLWKMLNIDVSATAFYKAQPVIEFCEVLDFKSEQOKPLTDSQVRFTKEIGLKYVE 272
Qy 224 TKY---NNKTYRVDIDMDGNPKSTFKKADGS-----EVSFLEYRKYQNOEITDLKQPV 276
Db 273 ITHCGQMKRYRVCNTRRPASHQTPPLQESGQTVECTVAQYFKDRHKLVLRYPHLPCL 332
Qy 277 -VSQPKRRRGGTLPAPMLIPELCVLTG-----LTDKMRDNFNMKDLAVHTRTP 328
Db 333 QVQEQKH-----TYLPLEVCNIVAGQCIKLTLD---NOTSTM--IRATARSAP 377
Qy 329 EQRQREVGKLIDYIHKNDNVQRELDRWGLSFDGNLLSFGRIILQTEKHGGKTDFYDQ 388
Db 378 D-RQBEISKLME--SASFNTDPYREFGIMVXDEMVDVTRGVLPQPPSILYGR-----427
Qy 389 PADWSKETGAPLISVKPLDN-----WLLIYTERNYEAA-----NSLIQNLFKYT 433
Db 428 -----NKAIATPVQGVDMRNKQFHTGIEIKVWATACFAPQCRQCTEVHLKSTEQRLKIS 482
Qy 434 PAMGQMRK-----AIMIEVDDRTYAYLRVLQOKVTADTQIVVCLSSNRKDKYDAIKKYL 489
Db 483 RADGMPIQGQPCFCKYAQAGSDSVEPMRHL-KNTYAGLQLVVIL-PGKTPVTAEVKRVG 540
Qy 490 CTDCPTPSQCVVARTLGKQOTVMAIATKIALQMNCKMGG-----ELMRVDPILKLVMTIV 543
Db 541 DTVLGNMATQCVQKNNV--QRTPPTQTLNCLKINVKLGGVNNILLPQGRPPVPFQPVIFL 598
Qy 544 GIDCVHDMTAG---RRSIAGFVASINEGTRMFRSCIFQDRGOELVDGLKVCLOALRAW 600

```

Db 599 GADVTHP-PAGDGGKPSIAAVGVSMADHPNRYCATVRVQQHREIIQDLAAMVRELLIQF 657
 Qy 601 NSCENYMPRIIIVRDCVGGQKTLVNVVDFDCLXSGRGYNPRLTVIWKXVNT 660
 Db 658 YKSTFRKPTRIIFYRDGVSEGGFQQVHLHELLAIREACIKLEKEYQPGITFIVQKRHT 717
 Qy 661 RFPA-----QSGRLQNPCTVIDVTRPEWDFPIVSOAVRSQSVSTHNVIVDNGS 716
 Db 718 RLFCCTKXNVRKSGNIPAGTIVDTKITHPTBEDFYLCSHAGIQGSRFSHVHLWDDNR 777
 Qy 717 LKPDHIQRLTYKLCHIYNMGPVIRPAPCOYAKLAF 754
 Db 778 FSSDEQLITLYQLCHYVRCRVSIPAPAYVHLVAF 815

RESULT 5
 PINH ARATH
 ID PINH ARATH STANDARD; PRT; 988 AA.
 AC Q9XGW1; O49256;
 DT 16-OCT-2001 (Rel. 40, Created)
 DT 16-OCT-2001 (Rel. 40, Last sequence update)
 DT 28-FEB-2003 (Rel. 41, Last annotation update)
 DE PINHEAD protein (ZWILLE protein).
 GN AT5G43810 OR MOD19.17.
 OS Arabidopsis thaliana (Mouse-ear cress).
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
 OC eurosids II; Brassicales; Brassicaceae; Arabidopsids.
 OX NCBI_TaxId=3702;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=cv. Landsberg erecta;
 RX MEDLINE=99094873; PubMed=9876176;
 RA Lynn K.; Fernandez A.; Aida M.; Sedbrook J.; Tasaka M.; Masson P.;
 RA Barton M.K.;
 RT "The PINHEAD/ZWILLE gene acts pleiotropically in Arabidopsis
 RT development and has overlapping functions with the ARGONAUTE1 gene.";
 RL Development 126:469-481(1999).
 RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN=cv. Landsberg erecta;
 RX MEDLINE=98169396; PubMed=9501101;
 RA Moussian B.; Schoof H.; Haecker A.; Juergens G.; Laux T.;
 RT "Role of the ZWILLE gene in the regulation of central shoot meristem
 RT cell fate during Arabidopsis embryogenesis.";
 RL EMBO J. 17:1799-1809(1998).
 RN [3]
 RP SEQUENCE FROM N.A.
 RC STRAIN=cv. Columbia;
 RA Kaneko T.; Katoh T.; Asamizu E.; Sato S.; Nakamura Y.; Kotani H.;
 RA Tabata S.;
 RT "Structural analysis of Arabidopsis thaliana chromosome 5. XI.";
 RL Submitted (APR-1999) to the EMBL/GenBank/DBJ databases.
 CC -!- FUNCTION: DEVELOPMENTAL PROTEIN REQUIRED FOR RELIABLE FORMATION OF
 CC PRIMARY AND AXILLARY SHOOT APICAL MERISTEMS. MAY BE A COMPONENT OF
 CC A HYPOTHETICAL MERISTEM FORMING COMPETENCE FACTOR.
 CC -!- SIMILARITY: BELONGS TO THE ARGONAUTE FAMILY.
 CC -!- SIMILARITY: Contains 1 PZ domain.
 CC -!- SIMILARITY: Contains 1 PZ domain.
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@sib-sib.ch).
 CC -----
 DR ENBL; AF154272; AAD40098.1; -.
 DR ENBL; AJ223508; CAAL1429.1; -.
 DR ENBL; AB026651; BAB11310.1; -.
 DR PIR; T52134; T52134.
 DR InterPro; IPR003100; PZ.

DR InterPro; IPR003165; P1wi.
 DR Pfam; PF02170; PZ; 1.
 DR Pfam; PF02171; P1wi; 1.
 DR PROSITE; PS50821; PZ; 1.
 DR PROSITE; PS50822; P1wi; 1.
 KW Developmental protein.
 FT DOMAIN 337 451
 FT P1wi.
 FT DOMAIN 625 946
 FT CONFLICT 475 479
 FT CONFLICT 671 671
 FT N -> D (IN REF. 2).
 SQ SEQUENCE 988 AA; 110867 MW; 32EBB349C613DA20 CRC64;

Query Match 12.3%; Score 503; DB 1; Length 988;
 Best Local Similarity 23.7%; Pred. No. 1,4e-28;
 Matches 204; Conservative 136; Mismatches 399; Indels 122; Gaps 32;

Qy 5 VNTRQLDHVKESKTGSSGIIIVRLSTNHFRLTSRFPQWALYQYHIDYNPLMEARRLSAL- 63
 Db 120 MGVKNSNFAPRPGFGTLGTCIVKANHF-LADLPKDLNQYDVTITPEVSSKSVNRAII 178
 Qy 64 -----LQHEDLIGKHAFD-----GTLFLPKRLQOKVTE-----VFSKTRNGEDVR 106
 Db 179 AELVRLYKESDLGRRLPAYDGRKSLYTAGELPFTWKFSVKITVDEDDGLINGPKERSYK 238
 Qy 107 ITIT-----LTNELPPTSPTCLQFYNIIFRLLKIMNLQQICGRNVNPNDEPI 153
 Db 239 VAKFVARANMHLGEFLAGKADCPQAVQLDVLRE-LSVKRCPVGRSFFSPD--I 295
 Qy 154 DIP---SHRLVIWPGFTTSIIQYENSIMLCTDV-SHKVLRSETVLDPMFNFVHQTEHK- 208
 Db 296 KTFORLGELESWCIFYQSIRPTQGLSLNIDMASAAPIEPLPVIETFAQLLGKDVLSKP 355
 Qy 209 ----PQEQVSKELIGLVLTKN---NKTRYDDIDMDQNPSTKTKADGSEV-SLEYK 260
 Db 356 LSDSDRVKIKKGLRGVKEVTHRANVRKRYVAGLUTQFTRELMFPVDENCTMKSVEIF 415
 Qy 261 RKQYNQEIITDLKQPVL-VSQPKRRRPGGTLPGPAMLIPELCYLTGLTDMKDNDFNMKD 319
 Db 416 QEMYGFTQHTLPLCLOVGNQK---ASYLPNEACKIVEGORYT----KRLNEKQITAL 467
 Qy 320 LAVHTRLTPEQOREVGRLLIDYIHRNDNVQRELDRWGLSFDNSNLSFSGRIIOTE--KTH 377
 Db 468 LKV-TCQPRDRENDILRTVQ--HNAYDQDPYAKEFGMNISEKLASVEARILPAPWLKYH 524
 Qy 378 QSGKTFDYNQPADMSKETRGAPLISVKPLDMLLIYTERNYE--AANSLIQNLKFKVTA 435
 Db 525 ENGKXKDCLPQVGQNMNMNK--KMINGMTVSRWACVNFSSVQENVARFCNELGQMCV 582
 Qy 436 MGXQWEKAIMIEV---DDRTE-----AYLRVLQKQVTDATQIVVCLSSNRKDKYDAIK 486
 Db 583 SCWENPPEVPIYSARPDPQEKALKHVYHTSMNKTGKLELLAILPNNNGSLXGDLK 642
 Qy 487 KYLCTDCPTSPQCVAR---TLGKQQTVMATKIALQNCXNGGE-----LMRV-- 533
 Db 643 RICETELGLISQCLTKHVFKISKQ-----YLANVSLKINVRMGGRNTVLVDAISRIPL 697
 Qy 534 --DIPKLVMIVGIDCYHDMTA--GRRSTAGFVASIN-EGMTWFSRCIFQDGRGQELVDG 588
 Db 698 VSDIP---TIIFGADVTHFENGEESSPSIAAVASQDWPEVTYAGLVCAQAHQRLIQD 754
 Qy 589 LKVLQQAALR-----AWNSCNEYMPRIIVYRDGVGQQLKTLVNYEVPQL 635
 Db 755 LYKTWQDPVRGTVSGMIDLLISFRKATQKQPLAIFVRDGVSEGFQVLLYELDAIR 814
 Qy 636 DCLKSIGRGYNPRLTVIVVKRNVNTRFPA-----QSGRLQNPFGTVIVDVEVTRPEW 689
 Db 815 KACASLEPNYQPEVTFIVVQKSHHTLFPANNHRDKNSTDRSGNILLPGTVVDTKICHTPEF 874
 Qy 690 DFFIVSQAVRSQSVSPHYNVINDNSGLKAPDHIQRLTYKLCHIYNWNPVIRVPACQYA 749
 Db 875 DFYLCSHAGIQGTSRPAHYHVLWDENNFTADGIQSLTNLNCYTYARCTRSVSVIPPAYYA 934
 Qy 750 HKLAFIVGQSIHREPNIISLN 770

```

Db 935 HLAAPRA--RFVLEPEIMQDN 953

RESULT 6
AGOL ARATH STANDARD; PRT; 997 AA.
AC Q9SUK3;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Argonaute-like protein A2927880.
OS A2927880 OR TIE2.20.
GN Arabidopsis thaliana (Mouse-ear cress).
OC Eukaryota; Viridiplantae; Streptophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
OC eucosids II; Brassicales; Brassicaceae; Arabidopsi.
OC NCBI_TaxID=3702;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=cv. Columbia;
RX MEDLINE=20083487; PubMed=10617197;
RA Lin X., Kaul S., Rounsley S.D., Shea T.P., Benito M.-I., Town C.D.,
RA Fujii C.Y., Mason T.M., Bowman C.L., Barnstead M.E., Feldblyum T.V.,
RA Buell C.R., Ketchum K.A., Lee J.J., Rensing C.M., Koo H.L.,
RA Moffat K.S., Cronin L.A., Shen M., Pal G., Van Aken S., Umayam L.,
RA Tallon L.J., Gill J.E., Adams M.D., Carrera A.J., Creasy T.H.,
RA Goodman H.M., Somerville C.R., Copenhaver G.P., Preuss D.,
RA Nickman W.C., White O., Bisen J.A., Salzberg S.L., Fraser C.M.,
RA Venter J.C.;
RT "Sequence and analysis of chromosome 2 of the plant Arabidopsis
RT thaliana."
RL Nature 402:761-768 (1999).
CC -!- SIMILARITY: BELONGS TO THE ARGONAUTE FAMILY.
CC -!- SIMILARITY: Contains 1 PAZ domain.
CC -!- SIMILARITY: Contains 1 Piwi domain.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; AC006929; AAD21514.1;
CC PIR; A84678; A84678.
CC InterPro; IPR003100; PAZ.
CC InterPro; IPR003185; Piwi.
CC Pfam; PF02170; PAZ; 1.
CC Pfam; PF02171; Piwi; 1.
CC PROSITE; P850821; PAZ; 1.
CC PROSITE; P850822; PIWI; 1.
CC FT DOMAIN 359 471 PAZ.
CC FT DOMAIN 638 958 PIWI.
CC SEQUENCE 997 AA; 111134 MW; DA93A1446C42F31 CRC64;

Query Match 12.1%; Score 495.5; DB 1; Length 997;
Best Local Similarity 24.1%; Pred. No. 5.1e-28;
Matches 200; Conservative 139; Mismatches 363; Indels 129; Gaps 30;

QY 20 GSGGIIVRLSTNHFRTSPQWALYQHIDYNFLMEA-----RLRSALLFQHEH--LICK 73
Db 154 GTLGGKVMVRANHF-LVQVADRLYHYDVSINDEVISKTVMNRVWMLKVNKDKSHLGK 212
QY 74 CHAFD-----GTLFLPKRLQKQVTFVFSKTRNGEDVRITITNELPTSPCLQF 125
Db 213 SPAYDGRKSLYTAGLPFFDSKEFVNLAERKDGSGKORPPKVAVN---VISTDLYQL 269
QY 126 YNIIFRR-----LLKINWL-----QQIGENYNPNDEIPSHR-----LVI 162
Db 270 QQFLDKRQREAPYDTIQVLVDVLRDPSNDYVSGSFHTSLGDKDARGELGDIY 329

```

```

163 WPGFTTSLQYENSIMLCTDVSHKVLRSSETVL-DFMNFNYQTBEHK-----FQGVSKSE 216
Db 330 WRGYFQSLRLTQMGLSLNIDVSARSFYEPVIVTDFISKFLNIRDNLRLPLRDSRLKVKV 389
QY 217 LIGL-VLITKYN-NKTYRVDDIDWDONPKSPFKADSGSEVSEFLEYRQYNOSEITDLKOP 274
Db 390 LRTLKVLKLLHWNCTKSAKISSIPRELFTLEDKSEKTVVQFAEKNYRVKYQALP 449
QY 275 VLVSQPKRRRGPGGTLPGFAMLIPELC-----YLTGLTDKMRNDENVMDLAVHRLT 327
Db 450 AIQT-----GSDTRPVVLPMLCQIDSGQRTYKELNEKQ-----VTALLKATCOR 494
QY 328 PEQOREVGRLLIDYIHKNDNVQRELRLDWGLSFDNSLLSFGSRIQLQTE--KIHGGKTFDY 385
Db 495 PPDRNSIKNLVKNVKNYNDLSKE---FGMSVTTQLASIEARVLPPPMKLVHDSKERNV 551
QY 386 NQOFADWS---KETRGAPLISVKEPLDNWLLI--YTRRYEAAANSLLIQLNFKVTFAMNQ 439
Db 552 NPLRGQMNNDKQWNGAKVTS-----WTCVSFSTRIDRGLPQBEFCQKOLIGMCVSKGME 605
QY 440 MRKA-----IMIEVDDRTREAYLRVLQKVTADTQIVVCLLSSNRKDKYDAIKKYLCTDCP 494
Db 606 PKQPAIFPISCPPEHIEALLDI--HKRAGLQLLIVIL-PDVTGSYKIKRICETELG 662
QY 495 TPSQCVARTIGKQQTVMATKIALQNCXKGG-----ELMRVDIPL---KLVNIGVI 545
Db 663 IVSQCCQPRQVKNL--KQYMNVAIKNVKTGGRNTVLNDAIRRNIPILITDRPTIIMGA 720
QY 546 DCYHDMTA--GERSTAGFVASIN-EGMTFWPSCRIFDQGOELVDGLKVCLOAAR--- 598
Db 721 DVTHPQGEDSPSIAAVASMDWPEINKYRGLVSAQAHRHEIIOQLKVLVODPORGLVH 780
QY 599 -----AMNSCNEYMPRIIVRDGVGQGLKTLVNYEVPQFELDLCKLSIGRGNPLR 649
Db 781 SGLIREHPIARRAQIPQRIIFRDGVSEGPSQVLLHMTAARKACNSLQENYVPRV 840
QY 650 TVIVVKKVNTRFFAQSOG-----RLQNPIPGTVIDVEVTRPEWYDFIVSQAVRSGSV 703
Db 841 TFVIVQKRHHTLPEQHGNDMTKSGNIQGTVDVTKICHNEFEFDVFLNSHAGTQGS 900
QY 704 SPTHVNIYDNLGKPDHIOBLTYKLCHLYNWPQVIRVPAPQYAHKLAF 754.
Db 901 RPAHYHLLDENGFTADQLQMLTNLCYARCTKSIVPPPAYVLAALAF 951

RESULT 7
I2C2 RABIT
ID I2C2_RABIT STANDARD; PRT; 813 AA.
AC O77503;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Eukaryotic translation initiation factor 2C 2 (eIF2C 2) (eIF-2C 2).
GN EIF2C2.
OS Oryctolagus cuniculus (Rabbit).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Lagomorpha; Leporidae; Oryctolagus.
OC NCBI_TaxID=9986;
RN [1]
RP SEQUENCE FROM N.A., AND PARTIAL SEQUENCE.
RC TISSUE=Liver;
RX MEDLINE=98267198; PubMed=9602122;
RA Zou C., Zhang Z., Wu S., Osterman J.C.;
RT "Molecular cloning and characterization of a rabbit eIF2C protein.";
RL Gene 211:187-194 (1998).
CC -!- FUNCTION: Plays an important role in the eukaryotic peptide chain
CC initiation process.
CC -!- SIMILARITY: BELONGS TO THE ARGONAUTE FAMILY.
CC -!- SIMILARITY: Contains 1 PAZ domain.
CC -!- SIMILARITY: Contains 1 Piwi domain.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -

```

D_b 687 GTTVDTKITHPTFFDYLGSHAGIQTSRPSHYHVLWDDNRPSSDELQILTYQLCHTYVR 746

QY 736 WPGVIRVPAPCOYAHKLAF 754

D_b 747 CTRSVSIPAPAYAHVLVAF 765

```

RESULT 8
YO43_CABEEL STANDARD; PRT; 1040 AA.
ID ID YO43_CABEEL
AC P34681;
DT 01-FEB-1994 (Rel. 28, Created)
DT 01-FEB-1995 (Rel. 31, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Hypothetical protein ZK757.3 in chromosome III.
GS ZK757.3.
OS Caenorhabditis elegans.
OC Eukaryota; Metazoa; Nematozoa; Chromadorea; Rhabditida; Rhabditoidea;
OC Rhabditidae; Pelodierinae; Caenorhabditis.
OX NCBI_TaxID=6239;
RN [1]
RS
SEQUENCE FROM N.A.
RP
RC STRAIN=Bristol N2;
RC MEDLINE=94150718; PubMed=7906398;
RX
RA Wilson R., Ainscough R., Anderson K., Baynes C., Berks M.,
RA Bonfield J., Burton J., Connell M., Copsey T., Cooper J., Coulson A.,
RA Craxton M., Dear S., Du Z., Durbin R., Favello A., Fraser A.,
RA Fulton L., Gardner A., Green P., Hawkins T., Hillier L., Jier M.,
RA Johnston L., Jones M., Kershaw J., Kirsten J., Laister N.,
RA Letellier P., Lightning J., Lloyd C., Mortimore B., O'Callaghan M.,
RA Parsons J., Percy C., Rifkin L., Roopra A., Saunders D., Showkeen R.,
RA Sims M., Smaldon N., Smith A., Smith M., Sonhammer E., Staden R.,
RA Sulston J., Thierry-Mieg J., Thomas K., Vaudin M., Vaughan K.,
RA Waterston R., Watson A., Weinstock L., Wilkinson-Sproat J.,
RA Wohlman P.;
RA
RT "2.2 Mb of contiguous nucleotide sequence from chromosome III of C.
RT elegans.";
RL Nature 368:32-38(1994).
CC CC -1- SIMILARITY: Contains 1 PAZ domain.
CC CC -1- SIMILARITY: Contains 1 Piwi domain.
CC CC -----
CC CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR ENBL; Z30215; CAA82941.1; -
DR ENBL; Z29121; CAA82941.1; JOINED.
DR ENBL; Z29121; CAA82389.1; -
DR ENBL; Z30215; CAA82389.1; JOINED.
DR PIR; D88568; D88568.
DR WormPep; ZK757.3; CE01117.
DR InterPro; IPR003100; PAZ.
DR InterPro; IPR003185; Piwi.
DR Pfam; PF02170; PAZ; 1.
DR Pfam; PF02171; Piwi; 1.
DR PROSITE; PS50821; PAZ; 1.
DR PROSITE; PS50822; PIWI; 1.
KW Hypothetical protein.
FT DOMAIN 378 486 PAZ.
FT DOMAIN 660 966 PIWI.
SQ SEQUENCE 1040 AA; 115415 MW; 4C8483C1FD72338 CRC64;
Query Match 10.3%; Score 421; DB 1; Length 1040;
Best Local Similarity 21.6%; Pred. No. 1.4e-22;
Matches 199; Conservative 148; Mismatches 367; Indels 206; Gaps 36;
QY 20 GSGGIVRLSTNHRL-TSRPQWALYQYHID-YNPLMEARRL---RSALLF-----CHED 69
Db 93 GTIGROIYKSNFFAMDLNKPKMWVIOVEVTHHP--GCKLKDKDWRIFKWKAVDHPN 150

```

```
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN SEQUENCE FROM N.A.
RP MEDLINE=20005943; PubMed=10534406;
RX Koesters R., Adams V., Betts D., Moos R., Schmid M., Siermann A.,
RA Hassam S., Weitz S., Lichter P., Heitz P.U., von Knebel Doeberitz M.,
RA Briner J.;
RT "Human eukaryotic initiation factor EIF2C1 gene: cDNA sequence,
genomic organization, localization to chromosomal bands 1p34-p35, and
expression.";
RL Genomics 61:210-218(1999).
CC -!- FUNCTION: Plays an important role in the eukaryotic peptide chain
initiation process.
CC -!- SIMILARITY: BELONGS TO THE ARGONAUTE FAMILY.
CC -!- SIMILARITY: Contains 1 Piwi domain.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (see http://www.isb-sib.ch/announce/
or send an email to license@isb-sib.ch).
CC -----
EMBL; AF121255; AAF13034.2; -
DR Genew; HGNC:3263; EIP2C2.
DR GO; GO:0003743; P:translation initiation factor activity; NAS.
DR GO; GO:0006412; P:protein biosynthesis; NAS.
DR InterPro; IPR003165; Piwi.
DR Pfam; PF02171; Piwi; 1.
DR PROSITE; PS0822; PIWI; 1.
KW Initiation factor; Protein biosynthesis.
FT NON_TER 1
FT DOMAIN 35 336 PIWI.
FT SEQUENCE 377 AA; 42529 MW; 3BCE91689D58526 CRC64;
SQ
Query Match 9.1%; Score 374; DB 1; Length 377;
Best Local Similarity 30.2%; Pred. No. 8.5e-20;
Matches 96; Conservative 64; Mismatches 140; Indels 18; Gaps 7;
QY 450 DRTAYLRVLOQKVADTQIVVCLSSNRKDKDAIKYLCDCPTPQCVVARTLGKQQ 509
Db 17 DSVEPMFRHL-KNTYAGLQVVL-PGKTPVYAEVKRVGDTVLGMATQCVQMKNV--QR 72
QY 510 TVMATATKALQNMCKMG-----ELMRVDIPLKLVMIIVGIDCYHDMTAG---RESIAG 560
Db 73 TTFQTLNMLKINVKLGGMNILLPQGRPPVFBEPVIFLGADVTHP-PAGDKKPSIAA 131
QY 561 FVASINEGTRWFRCIFQDRGOELVDGLKVCLOALRAWNSCNEYMPRIIVYRDGVD 620
Db 132 VVGRMDAHPNRYCATRVQQRHEIIQDLAAMVRELLIQFYKSTRFKPRIIFYRDGVSE 191
QY 621 GQLKTLVNVVPOFLDLCKSIGRGNPRILTVIVVKRVNTRFFA-----QSGRLQNPLEG 676
Db 192 GQFQVLLHLLAIAREACIKLEKDYQPGITFTVWQKRHHRLFCCTDKNERVKGSGNIPAG 251
QY 677 TVIDVEVTRPEWYDFIVSQAVERSGSVPTHYNVLYDNSGLKPKDHIQRLTYKLCHIYNW 736
Db 252 TVVDYKITHPTFEFDYLCSHAGIQGTSRPSHYHLVLDNDRFSSDELQILTYQLCHTYVRC 311
QY 737 PGVIRVPAPCOYAHKIAF 754
Db 312 TRSVSIPAPAYVAHLVAF 329
RESULT 10
Y053 CAEEL
ID Y053 CAEEL STANDARD; PRT; 1032 AA.
AC Q09249;
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
```

```
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN SEQUENCE FROM N.A.
RP MEDLINE=20005943; PubMed=10534406;
RX Koesters R., Adams V., Betts D., Moos R., Schmid M., Siermann A.,
RA Hassam S., Weitz S., Lichter P., Heitz P.U., von Knebel Doeberitz M.,
RA Briner J.;
RT "Human eukaryotic initiation factor EIF2C1 gene: cDNA sequence,
genomic organization, localization to chromosomal bands 1p34-p35, and
expression.";
RL Genomics 61:210-218(1999).
CC -!- FUNCTION: Plays an important role in the eukaryotic peptide chain
initiation process.
CC -!- SIMILARITY: BELONGS TO THE ARGONAUTE FAMILY.
CC -!- SIMILARITY: Contains 1 Piwi domain.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (see http://www.isb-sib.ch/announce/
or send an email to license@isb-sib.ch).
CC -----
EMBL; AF121255; AAF13034.2; -
DR Genew; HGNC:3263; EIP2C2.
DR GO; GO:0003743; P:translation initiation factor activity; NAS.
DR GO; GO:0006412; P:protein biosynthesis; NAS.
DR InterPro; IPR003165; Piwi.
DR Pfam; PF02171; Piwi; 1.
DR PROSITE; PS0822; PIWI; 1.
KW Initiation factor; Protein biosynthesis.
FT NON_TER 1
FT DOMAIN 35 336 PIWI.
FT SEQUENCE 377 AA; 42529 MW; 3BCE91689D58526 CRC64;
SQ
Query Match 9.1%; Score 374; DB 1; Length 377;
Best Local Similarity 30.2%; Pred. No. 8.5e-20;
Matches 96; Conservative 64; Mismatches 140; Indels 18; Gaps 7;
QY 450 DRTAYLRVLOQKVADTQIVVCLSSNRKDKDAIKYLCDCPTPQCVVARTLGKQQ 509
Db 17 DSVEPMFRHL-KNTYAGLQVVL-PGKTPVYAEVKRVGDTVLGMATQCVQMKNV--QR 72
QY 510 TVMATATKALQNMCKMG-----ELMRVDIPLKLVMIIVGIDCYHDMTAG---RESIAG 560
Db 73 TTFQTLNMLKINVKLGGMNILLPQGRPPVFBEPVIFLGADVTHP-PAGDKKPSIAA 131
QY 561 FVASINEGTRWFRCIFQDRGOELVDGLKVCLOALRAWNSCNEYMPRIIVYRDGVD 620
Db 132 VVGRMDAHPNRYCATRVQQRHEIIQDLAAMVRELLIQFYKSTRFKPRIIFYRDGVSE 191
QY 621 GQLKTLVNVVPOFLDLCKSIGRGNPRILTVIVVKRVNTRFFA-----QSGRLQNPLEG 676
Db 192 GQFQVLLHLLAIAREACIKLEKDYQPGITFTVWQKRHHRLFCCTDKNERVKGSGNIPAG 251
QY 677 TVIDVEVTRPEWYDFIVSQAVERSGSVPTHYNVLYDNSGLKPKDHIQRLTYKLCHIYNW 736
Db 252 TVVDYKITHPTFEFDYLCSHAGIQGTSRPSHYHLVLDNDRFSSDELQILTYQLCHTYVRC 311
QY 737 PGVIRVPAPCOYAHKIAF 754
Db 312 TRSVSIPAPAYVAHLVAF 329
RESULT 10
Y053 CAEEL
ID Y053 CAEEL STANDARD; PRT; 1032 AA.
AC Q09249;
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
```

DT 28-FEB-2003 (Rel. 41, Last annotation update)
 DE Hypothetical protein C16C10.3 in chromosome III.
 GN C16C10.3.
 OS Caenorhabditis elegans.
 CC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditidae;
 CC Rhabditidae; Peloderinae; Caenorhabditis.
 OX NCBI_TaxID=6239;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=Bristol N2;
 RA Lloyd C.;
 RL Submitted (NOV-1994) to the EMBL/GenBank/DBJ databases.
 CC -!- SIMILARITY: STRONG, TO C.ELEGANS C14B1.7. ALSO TO C06A1.4,
 CC F49F7.1, R06C7.1, T22B3.2 AND ZK757.3.
 CC -!- SIMILARITY: Contains 1 PAZ domain.
 CC -!- SIMILARITY: Contains 1 Piwi domain.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL; 246787; CAA86741.1; --
 CC PIR; T19324; T19324.
 CC WormPep; C16C10.3; CB01494.
 CC InterPro; IPR003100; PAZ.
 CC Pfam; PF02170; PAZ; 1.
 CC Pfam; PF02171; Piwi; 1.
 CC PROSITE; PS50821; PAZ; 1.
 CC PROSITE; PS50822; PIWI; 1.
 CC Hypothetical protein.
 KW DOMAIN 369 481
 FT DOMAIN 650 977
 FT PIWI.
 SQ SEQUENCE 1032 AA; 117105 MW; EBEAE23CE67BB149 CRC64;

 Query Match 6.1%; Score 248.5; DB 1; Length 1032;
 Best Local Similarity 20.9%; Pred. No. 4.9e-10;
 Matches 131; Conservative 98; Mismatches 229; Indels 169; Gaps 24;

 QY 206 EHKFOEQ-----VSKELIGL-VLTLYNNKTYRVDIDWDQNPXSTFKKADGS---EVS 255
 DB 381 EHAFGERNKTCIEBALKGLDVECTHLKGNLRVSSIAENNAENTSFMKDKDGERVT 440
 QY 256 FLEYTRKQVNEITDLKQVLVSQPKR-----RKGPGGTLPGPAMLIPELCYLTG 305
 DB 441 VAEYFLLQYNIKLKYPRLPLVWVKRPFKESFFPMELLRIAPQRIKVNKNSPTVQSAMTG 500
 QY 306 LTDKYRDNFNMKDLAVHTRLTPEQORQEVGLIDYIHKNDNVQELRDWGLSFDNLLS 365
 DB 501 RNASH-----PQHVKLVQDIL-----RDNLKLEQNKYMDAPFGIKLMS 538
 QY 366 -----FSGRILQTEKHGGKTF--DYN--PQADWSKETRGAPL--ISVKPLDNLILI-- 413
 DB 539 TEPIQWATKLPPAIIKFGQTYMDMSEPAFTQDKFVEPARIRKIGIIVFDNCIQMRQ 598
 QY 414 -----YTRNYEAANSL-IQNLFKVTTPAMGMQMRKAIMIE 447
 DB 599 ABDFCDKLSNFCRDNGITVEKSDRWSIRELNSDSVAIONLMKKW----- 644
 QY 448 VDDRTEAYLRVLOQKVATDTQIVCLLSNRKDKYDAIKY-----LCTDCTPS 497
 DB 645 LDDRVDILVGIAREK-----KPDVHDILKYFESIGLQTIQLC----- 682
 QY 498 QCVWARTLKGQ---QTVMAITKIALQMNCKMGGLWRVDIP----- 536
 DB 683 QQTVDKMGCGGQRTIDNMVK-----FNLKCGGNFFVEIENAVRGKAVCSNNTLRKK 738
 QY 537 -LKLVMIVGIDCYHDMTA-----GRSIAAGFVASINEGMTRWFSRCIFQDRGOE 584

RESULT 11

ID DICE HUMAN STANDARD; PRT; 1912 AA.
 AC Q9UPY3; O95943; Q9U002;
 DT 28-FEB-2003 (Rel. 41, Created)
 DT 28-FEB-2003 (Rel. 41, Last sequence update)
 DE Endoribonuclease Dicer (EC 3.1.26.-) (Helicase with RNase motif)
 DE (Helicase-MOI).
 GN DICER1 OR DICER OR HERNA OR KIAA0928.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 OX NCBI_TaxID=9606;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=20246304; PubMed=10786632;
 RA Matsuda S., Ichigotani Y., Okuda T., Irimura T., Nakatsugawa S.,
 RA Hamaguchi M.;
 RT "Molecular cloning and characterization of a novel human gene (HERNA)
 RT which encodes a putative RNA-helicase.";
 RL Biochim. Biophys. Acta 1490:163-169(2000).
 RN [2]
 RP SEQUENCE FROM N.A.
 RC TISSUE=Lung;
 RA Provost P., Dishart D., Doucet D., Hermansson A., Frensdewey D.,
 RA Samuelsson B., Radmark O.;
 RT "RNA binding and processing by recombinant human Dicer.";
 RN Submitted (JUN-2002) to the EMBL/GenBank/DBJ databases.
 RN [3]
 RP SEQUENCE FROM N.A.
 RC TISSUE=Brain;
 RX MEDLINE=99246063; PubMed=10231032;
 RA Nagase T., Ishikawa K.-I., Suyama M., Kikuno R., Hirosewa M.,
 RA Miyajima N., Tanaka A., Kotani H., Nomura N., Ohara O.;
 RT "Prediction of the coding sequences of unidentified human genes. XIII.
 RT The complete sequences of 100 new cDNA clones from brain which code
 RT for large proteins in vitro.";
 RL DNA Res. 6:63-70(1999).
 RN [4]
 RP REVISIONS.
 RX MEDLINE=22158633; PubMed=12168954;
 RA Nakajima D., Okazaki N., Yamakawa H., Kikuno R., Ohara O., Nagase T.;
 RT "Construction of expression-ready cDNA clones for KIAA genes: manual
 RT curation of 330 KIAA cDNA clones.";
 RL DNA Res. 9:99-106(2002).
 RN [5]
 RP SEQUENCE OF 1238-1912 FROM N.A.
 RC TISSUE=Lung;
 RX MEDLINE=99162526; PubMed=10051563;
 RA Provost P., Samuelsson B., Radmark O.;
 RT "Interaction of 5-lipoxygenase with cellular proteins.";
 RL Proc. Natl. Acad. Sci. U.S.A. 96:1881-1885(1999).
 CC -!- FUNCTION: Involved in cleaving double-stranded RNA in the RNA
 CC interference (RNAi) pathway. It produces 21 to 23 bp dsRNAs
 CC (siRNAs) which target the selective destruction of homologous
 CC RNAs.
 CC -!- SIMILARITY: BELONGS TO THE HELICASE FAMILY.

Db 739 LLEHVQFTGFIISHGASRTLFDRSRQMDGPSVVGVSVSLTNS-TQLGGFTYLTQKEY 797
 QY 585 LVDGLKVCLOALRAWNSCNHMPRIIVYRDGVDGQGLTLVNVVEVQFLDCLXISGRG 644
 Db 798 KLOKLDFFPFCVRSYKSHKSLTFLTRIVYRVGAGEGNFNR-VKEEVEEMRRTFDKIQPG 856
 QY 645 YNPLRTIIVWKKRVNTRFFAQ--SGRL--QNPLPGTVIDVEVTRPEWVDRIIVSQAVRS 700
 Db 857 YRPHLVIIAQRASHARVFPSCISGNRATDQNIPTGTCVENVLISGYDDEFILSSQTPLI 916
 QY 701 GSVSPHYNVIVDNSGLKPDHIORLTY 727
 Db 917 GTVRPCKYITLVNDAKWSKNELMHTY 943

QY	210	OSQVSKELIGL	-----VLTKYN-----KTYRVDDIDWDNQPKS	244
Db	893	-----SEARIGIPSTKYTKETPFVFKLEDYQDAVILIPRYRNPDQHRFYADVVTDLTPLS	948	
QY	245	TFKKADGSEVSFLYYRQYNOEITDLKQVIL	-----VSQ	279
Db	949	KFPSPSEYE--TFAEYKYTKYNLDLNLNQPLLDVDTSSRLMLLTPRHNLNQKGLPLSS	1006	
QY	280	PKRRRGPGGTLPGPAMLIPELCYLGLTDMKNDNFNMKDL--AVHTRLTPQROREVG	337	
Db	1007	AEKRAKWESLQNKQILYPCLCAIHPIPASLWKAVCLPSILYRLHCLLTAEELRAQTAS	1066	
QY	338	-----LIDYIHKNDNVORELDMGLSFDPS--NLLSPS-----GRILQTE	374	
Db	1067	DAGVGVRSPLPADFRYPNLDF-----GWKKSIDSFISINSSSAENDNYCKHSTIVPEN	1121	
QY	375	KTHGGKTFDYNPQADWSKETRGAPLLSVKP	-----LDNWL	412
Db	1122	AAHQGANRTSSLENHDSQVNCR--TLSESPGKLVHSADLTAINGLSYNQNLANGSY	1179	
QY	413	IYTRNEYAANSLIQNLFPK---VTPAMGMQMKAIMIEVDDR-----TEAIVRLVQ	460	
Db	1180	DLANRDFCGGNQL--NYYKQETFPVQPTTSYSTONLSYENQPSDEBCTLLSNKYLDGNA	1237	
QY	461	QKVYADTQIVVCLLSNRKDKYDAIKKYLCTDPTPSQCQVARTLGKQOTVMAIATKIAL	520	
Db	1238	NKSTSDGSPVMAVMPGT-TDITQVLKGRMDSE-QSPSICYSRTTLPNPLGLILQALTL	1294	
QY	521	QNNKMGGLGEV-----DIPLKLVMLVGDIC-YHDMTAGRRRIAGFVASINSGMTWFSR	575	
Db	1295	-----NADSGFNLERLEMGDSFLKHAITLYICTYPAHEGRLSYM-----	1336	
QY	576	CIFDRGQGLVDGLKVCQLQAALRAWNSCNEY-----MPSRIIV	613	
Db	1337	-----RSXKV-----SNCNLYRLGKKKGILPSEMVV	1361	

RESULT 12

DICE MOUSE

ID	DICE MOUSE	STANDARD;	PRT;	1906 AA.
AC	Q8R4T8;			
DT	28-FEB-2003 (Rel. 41, Created)			
DT	28-FEB-2003 (Rel. 41, Last sequence update)			
DT	28-FEB-2003 (Rel. 41, Last annotation update)			
DE	Endoribonuclease Dicer (EC 3.1.26.-) (Double-strand-specific			
DE	ribonuclease mDCR-1)			
GN	DICER1 OR DICER OR MDCR.			
OS	Mus musculus (Mouse).			
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;			
OC	Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.			
NCBI	TaxID=10090;			
RN	[1] _			
RP	SEQUENCE FROM N.A.			
RX	MEDLINE=21886641; PubMed=11889553;			
RT	Nicholson R.H.; Nicholson A.W.;			
RT	"Molecular characterization of a mouse cDNA encoding Dicer, a			
RL	ribonuclease III ortholog involved in RNA interference.";			
RL	Mamm. Genome 13:67-73(2002).			
RN	[2]			
RP	SEQUENCE FROM N.A.			
RC	STRAIN=Czech II;			
RA	Svoboda P., Anger M., Stein P., Schultz R.M.;			
RT	"Mouse dicer homolog in oocyte and preimplantation embryos.";			
RL	Submitted (OCT-2001) to the EMBL/GenBank/DBJ databases.			
RN	[3]			
RP	SEQUENCE FROM N.A.			
RA	Doi N., Zenko S., Ui-Tei K., Takahashi F., Ueda R., Miyata Y.,			
RA	Saigo K.;			
RT	"e12C family proteins and Dicer homologs are required for siRNA			
RL	mediated RNAi in mammalian cells.";			
RL	Submitted (MAR-2002) to the EMBL/GenBank/DBJ databases.			
CC	CC -! FUNCTION: Involved in cleaving double-stranded RNA in the RNA			
CC	interference (RNAi) pathway. Itroduces 21 to 23 bp dsRNA			

411 1111
-----YKKNIEAANSLIQ-----NLI-----FNVIFAMGM-QMKKAINLEVDRT 452

CC	-----	
DR	EMBL; L18881; AAA27636.1; -.	
DR	EMBL; Z35454; CAA84606.1; -.	

```

RESULT 14
HMP2_YEREN STANDARD; PRT; 2035 AA.
ID HMP2_YEREN STANDARD; PRT; 2035 AA.
AC P48633;
DT 01-FEB-1996 (Rel. 33, Created)
DT 01-FEB-1996 (Rel. 33, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE High-molecular-weight protein 2 (HMWP2).
GN IRP2.
OS Yersinia enterocolitica.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC Enterobacteriaceae; Yersinia.
OX NCBI_TaxID=630;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=8081 / Serotype O:8;
RX MEDLINE=93374844; PubMed=8366034;
RA Guilvout I., Mercereau-Puijalon O., Bonnefoy S., Pugsley A.P.,
RA Carniel B.;
RT "High-molecular-weight protein 2 of Yersinia enterocolitica is
RT homologous to Angr of Vibrio anguillarum and belongs to a family of
RT proteins involved in nonribosomal peptide synthesis.";
RL J. Bacteriol. 175:5488-5504(1993).
CC -1- FUNCTION: UNKNOWN. MAY BE INVOLVED IN THE NONRIBOSOMAL SYNTHESIS
CC OF SMALL PEPTIDES.
CC -1- COFACTOR: CONTAINS 3 COVALENTLY BOUND PHOSPHOPANTETHEINES
CC (POTENTIAL).
CC -1- DOMAIN: CONSISTS OF A CENTRAL REGION WITH SIMILARITY TO THE REPEAT
CC DOMAINS OF ACVS AND GRC2, FLANKED BY TWO REPEAT DOMAINS, EACH OF
CC WHICH CONTAINS 5 DIRECT REPEATS.
CC -1- SIMILARITY: BELONGS TO THE ATP-DEPENDENT AMP-BINDING ENZYME
CC FAMILY.
CC -1- SIMILARITY: Contains 3 acyl carrier domains.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.isb-sib.ch/announce/)
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; L18881; AAA27636.1; -
CC EMBL; Z35454; CAA84606.1; -
CC PIR; A48654; A48654.

```


GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: December 6, 2003, 17:51:05 ; Search time 4855 Seconds
(without alignments)
11654.134 Million cell updates/sec

Title: US-10-043-774B-1

Perfect score: 2328

Sequence: 1 atgatctttggtgtaacac.....accgcctttactacctctaa 2328

Scoring table: OLIGO_NUC

Gapop 60.0 , Gapext 60.0

Searched: 22781392 seqs, 12152238056 residues

Word size : 0

Total number of hits satisfying chosen parameters: 45562784

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database :

EST:*

- 1: em_estba:*
- 2: em_esthum:*
- 3: em_estin:*
- 4: em_estnu:*
- 5: em_estov:*
- 6: em_estpl:*
- 7: em_estro:*
- 8: em_hic:*
- 9: gb_est1:*
- 10: gb_est2:*
- 11: gb_hic:*
- 12: gb_est3:*
- 13: gb_est4:*
- 14: gb_est5:*
- 15: em_estfun:*
- 16: em_estom:*
- 17: em_gss_hum:*
- 18: em_gss_inv:*
- 19: em_gss_pln:*
- 20: em_gss_vit:*
- 21: em_gss_fun:*
- 22: em_gss_man:*
- 23: em_gss_mus:*
- 24: em_gss_pro:*
- 25: em_gss_rod:*
- 26: em_gss_pig:*
- 27: em_gss_vrl:*
- 28: gb_gss1:*
- 29: gb_gss2:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	454	19.5	527	12	BM751055
2	452	19.4	874	10	BF666315
3	452	19.4	1495	10	BF664941
4	413	17.7	413	9	AL041473

5	405	17.4	642	13	BU854142
6	361	15.5	587	9	AA430311
7	321	13.8	462	9	AA969938
8	318	13.7	779	12	BI560710
9	308	13.2	818	12	EG773137
10	285	12.2	622	9	AL705072
11	275	11.8	559	12	BM728018
12	257	11.0	420	13	BQ374910
13	249	10.7	718	13	BQ435882
14	245	10.5	512	12	EM680592
15	241	10.4	640	9	AL705301
16	241	10.4	1092	12	BI463482
17	232	10.0	378	10	BF764928
18	226	9.7	226	10	BF092403
19	223	9.6	731	12	BI458664
20	218	9.4	715	10	EG721550
21	215	9.2	581	10	EG718168
22	208	8.9	705	10	BF246708
23	195	8.4	226	10	BF092519
24	192	8.2	605	9	AL704986
25	179	7.7	539	12	BI463615
26	178	7.6	666	12	BI463897
27	164	7.0	807	13	BU160046
28	125	5.4	522	9	AA456921
29	123	5.3	882	12	BI459464
30	108	4.6	470	13	EX281582
31	99	4.3	474	9	AL705611
32	91	3.9	830	10	BF699368
33	83	3.6	533	10	BF080923
34	83	3.6	577	10	BF080918
35	75	3.2	806	10	EG724366
36	71	3.0	744	12	BI459142
37	68	2.9	101	10	BF092393
38	60	2.6	113	13	EQ374884
39	57	2.4	935	10	BF247005
40	54	2.3	228	13	BQ375179
41	41	1.8	485	10	EG718140
42	35	1.5	477	12	BM53361
43	35	1.5	565	10	EG101107
44	34	1.5	872	12	BI560401
45	21	0.9	348	13	BY018987

ALIGNMENTS

RESULT 1
BM751055
LOCUS
DEFINITION K-EST0026956 S9SNU601 Homo sapiens cDNA clone S9SNU601-21-E08 5', linear EST 04-MAR-2002
mRNA sequence.
ACCESSION BM751055
VERSION BM751055.1 GI:19080673
KEYWORDS EST.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 527)
AUTHORS Kim,N.S., Hahn,Y., Oh,J.H., Lee,J.Y., Ahn,H.Y., Chu,M.Y., Kim,M.R., Oh,K.J., Cheong,J.E., Sohn,H.Y., Kim,J.M., Park,H.S., Kim,S. and Kim,Y.S.
21C Frontier Korean EST Project 2001
Unpublished
Contact: Kim YS
Genome Research Center
Korea Research Institute of Bioscience & Biotechnology
52 Eoeun-dong Yuseong-gu, Daejeon 305-333, South Korea
Tel: +82-42-860-4470
Fax: +82-42-860-4409
Email: yongsung@mail.kribb.re.kr
Plate: 21 row: E column: 08
High quality sequence stop: 527.


```

|||||
Db 310 CGATTGTTACCATGATGACAGCTGGCGGAGTCAATCGCAGGATTTGTTGCCAGCAT 369
|||||
QY 1695 CAATGAGGATGACCCGCTGGTCTCAGCTGCATATTTACGATAGAGGACAGAGCT 1754
|||||
Db 370 CAATGAAGGATGACCCGCTGGTCTCAGCTGCATATTTACGATAGAGGACAGAGCT 429
|||||
QY 1755 GTAGATGGCTCAAAAGTCTGCTGCAAGCGG 1786
|||||
Db 430 GGTAGATGGCTCAAAAGTCTGCTGCAAGCGG 461
|||||

BF664941 1495 bp mRNA linear EST 21-DEC-2000
602119191F1 NIH_MGC_56 Homo sapiens cDNA clone IMAGE:4276247 5',
mRNA sequence.
BF664941
BF664941.1 GI:11938836
EST.
Homo sapiens (human)
Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
NIH-MGC http://mgs.nci.nih.gov/
National Institutes of Health, Mammalian Gene Collection (MGC)
Unpublished
Contact: Robert Strausberg, Ph.D.
Email: cgapbs@mail.nih.gov
Tissue Procurement: ATCC
CDNA Library Preparation: CLONETECH Laboratories, Inc.
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Incyte Genomics, Inc.
Clone Distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
Plate: LCM1096 row: f column: 24
High quality sequence stop: 511.
Location/Qualifiers
1..1495
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="IMAGE:4276247"
/tissue_type="primitive neuroectoderm"
/lab_host="DH10B (T1 phage-resistant)"
/clone_lib="NIH MGC 56"
/notes="Organ: Brain; Vector: pDNR-LIB (Clontech); Site_1:
SfiI (ggcgctcgcc); Site_2: SfiI (ggccattatggcc);
Double-stranded cDNA was prepared from cell line RNA. 5'
and 3' adaptors were used in cloning as follows: 5'
adaptor sequence: 5'-CAGGCCATTATGGCC-3' and 3' adaptor
sequence: 5'-ATTCTAGAGCGCGCGGACATG-dt(30)BN-3'
(where B = A, C, G and N = A, C, G, or T). Average
insert size 1.65 kb (range 0.9-4.0 kb). 15/15 colonies
contained inserts by PCR. This library was enriched for
full-length clones and was constructed by Clontech
Laboratories (Palo Alto, CA)."
```

BASE COUNT 476 a 280 c 536 g 203 t

ORIGIN

Query Match 19.4%; Score 452; DB 10; Length 1495;
 Best Local Similarity 100.0%; Pred. No. 5.9e-235;
 Matches 452; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

QY 1335 GATTGAAGTGATGACAGAACTCAAGCTACTTAAGATGCTTACAGCAAAAGTCAAGC 1394
|||||
Db 10 GATTGAAGTGATGACAGAACTCAAGCTACTTAAGATGCTTACAGCAAAAGTCAAGC 69
|||||
```

```

QY 1395 AGACACCCAGATAGTTGTCTGTGTTCAAGTAATCGAAGGACAAATACGATCTAT 1454
|||||
Db 70 AGACACCCAGATAGTTGTCTGTGTTCAAGTAATCGAAGGACAAATACGATCTAT 129
|||||
```

```

QY 1455 TAAAAATAACTGTGTACAGATTGGCCCTACCCCAAGTCAGTGTGTGTGGCCGACACTT 1514
|||||
Db 130 TAAAAATAACTGTGTGTACAGATTGGCCCTACCCCAAGTCAGTGTGTGTGGCCGACACTT 189
|||||
QY 1515 AGGCAAAACAGCAAACTGTTCATGGCCATTGCTACAAAGATTGGCCCTACAGATGAATGCAA 1574
|||||
Db 190 AGGCAAAACAGCAAACTGTTCATGGCCATTGCTACAAAGATTGGCCCTACAGATGAATGCAA 249
|||||
QY 1575 GATGGAGGAGAGCTCTCTGGAGGTGGACATCCCTCTGAAGCTCGTGTGATGATCGTTGGCAT 1634
|||||
Db 250 GATGGAGGAGAGCTCTCTGGAGGTGGACATCCCTCTGAAGCTCGTGTGATGATCGTTGGCAT 309
|||||
QY 1635 CGATTGTTACCATGACATGACAGCTGGCGGAGGTCAATCGCAGGATTTGTTGCCAGCAT 1694
|||||
Db 310 CGATTGTTACCATGACATGACAGCTGGCGGAGGTCAATCGCAGGATTTGTTGCCAGCAT 369
|||||
QY 1695 CAATGAAGGATGACCCGCTGGTCTCAGCTGCATATTTACGATAGAGGACAGGAGCT 1754
|||||
Db 370 CAATGAAGGATGACCCGCTGGTCTCAGCTGCATATTTACGATAGAGGACAGGAGCT 429
|||||
QY 1755 GTAGATGGCTCAAAAGTCTGCTGCAAGCGG 1786
|||||
Db 430 GGTAGATGGCTCAAAAGTCTGCTGCAAGCGG 461
|||||
```

RESULT 4
 AL041473/c

LOCUS

DEFINITION
 DKFZp434E2417.s1.434 (synonym: htes3) Homo sapiens cDNA clone
 DKFZp434E2417.3, mRNA sequence.

ACCESSION
 AL041473.1 GI:5420824

VERSION
 EST.

KEYWORDS
 SOURCE

ORGANISM
 Homo sapiens (human)

REFERENCE
 1 (bases 1 to 413)

AUTHORS
 Bloecker, H., Boecker, M., Brandt, P., Mewes, H.W., Gassenhuber, J. and
 Wiemann, S.

TITLE
 EST (Bloecker, et al.)

JOURNAL
 Unpublished

COMMENT
 Contact: Bloecker H

MPIS

Ingolstaedter Landstr.1, D-85764 Neuherberg, Germany

This is the 3' sequence of the clone insert

Clone from S. Wiemann, Molecular Genome Analysis, German Cancer

Research Center (DKFZ); Email s.wiemann@dkfz-heidelberg.de;

sequenced by GBF (National Research Centre for Biotechnology Ltd.,

Braunschweig/Germany) within the cDNA sequencing consortium of the

German Genome Project.

ri sequence also available.

This clone (DKFZp434E2417) is available at the RZPD in Berlin.

Please contact the RZPD: Ressourcenzentrum, Heubnerweg 6, 14059

Berlin-Charlottenburg, GERMANY; Email: clone@rzpd.de.

location/Qualifiers

1..413

/organism="Homo sapiens"

/mol_type="mRNA"

/db_xref="taxon:9606"

/clone="DKFZp434E2417"

/tissue_type="testis"

/dev_stage="adult"

/lab_host="DH10B"

/clone_lib="434 (synonym: htes3)"

/note="Vector: pSport1; Site_1: NotI; Site_2: SalI"

BASE COUNT 100 a 118 c 83 g 112 t

ORIGIN

Query Match 17.7%; Score 413; DB 9; Length 413;

Best Local Similarity 100.0%; Pred. No. 8.5e-214;

Matches 413; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1637 ATTGTTACCATGACATGACAGCTGGCGAGAGTCAATCGCAGGATTGTTCCCGACATCA 1696
 Db |||||
 QY 1697 ATGAAGGATGACCCGCTGGTTCACGCTGCATATTCAGATAGAGGACAGAGCTGG 1756
 Db |||||
 QY 1757 TAGATGGGCTCAAAAGTCTGCTCAAGCGGCTCTGAGGGCTTGAATAGCTGCAATGAGT 1816
 Db |||||
 QY 1817 ACATGCCCGCGGATCATGCTGACCGGATGCGGTAGAGAGCGCGCAGCTGAAAACAC 1876
 Db |||||
 QY 1877 TGGTGAACACTAGAGTGGCAGAGTCTTTGGATTGCTTAAATCCATTGGTAGAGTTACA 1936
 Db |||||
 QY 1937 ACCCTAGACTAACCGTAAATGCTGAGTGAAGAAAGAGTGAACACAGAGTTTTCCTCAGT 1996
 Db |||||
 QY 1997 CTGAGGAGAGCTTCAGATCCATCTCTGAAACAGTATTGATGTAGAGTT 2049
 Db |||||

RESULT 5
 BU854142
 LOCUS 642 bp mRNA linear EST 16-OCT-2002
 DEFINITION AGENCOURT_10402932 NIH_MGC_82 Homo sapiens cDNA clone IMAGE:6621314
 5', mRNA sequence.
 ACCESSION BU854142
 VERSION BU854142.1 GI:24039108
 KEYWORDS EST.
 SOURCE Homo sapiens (human)
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 REFERENCE 1 (bases 1 to 642)
 AUTHORS NIH-MGC <http://mgs.nci.nih.gov/>.
 TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
 JOURNAL Unpublished
 COMMENT Contact: Robert Strausberg, Ph.D.
 Email: cgabbs-remail.nih.gov
 Tissue Procurement: CLONTECH
 CDNA Library Preparation: CLONTECH Laboratories, Inc.
 DNA Sequencing by: The I.M.A.G.E. Consortium (LLNL)
 Cloned by: Agencourt Bioscience Corporation
 Cloning strategy: MGC clone distribution information can be
 found through the I.M.A.G.E. Consortium/LLNL at:
<http://image.llnl.gov>
 Plate: MCM2872 row: n column: 02
 High quality sequence stop: 360.

FEATURES
 source
 1. .642
 Location/Qualifiers
 /organism="Homo sapiens"
 /mol_type="mRNA"
 /db_xref="taxon:9606"
 /clone="IMAGE:6621314"
 /lab_host="DH10B (T1 phage-resistant)"
 /clone_lib="NIH_MGC_82"
 /note="Organ: testis; Vector: pDNR-LIB (Clontech); Site: 1:
 SfiI (ggcgctcgcc); Site 2: SfiI (ggcattatggcc); 5' and
 3' adaptors were used in cloning as follows: 5' adaptor
 sequence: 5'-CACGGCCATATGGCC-3' and 3' adaptor sequence:
 5'-ATTCTAGAGCCGAGCGCGGCATG-dt(30)BN-3' (where B = A,
 C, or G and N = A, C, G, or T). Average insert size
 1.35 kb (range 0.9-4.0 kb). 14/15 colonies contained
 inserts by PCR. This library was enriched for full-length

clones and was constructed by Clontech Laboratories (Palo
 Alto, CA)."
 BASE COUNT 194 a 138 c 177 g 133 t
 ORIGIN
 Query Match 17.4%; Score 405; DB 13; Length 642;
 Best Local Similarity 100.0%; Pred. No. 2.3e-209;
 Matches 405; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 1450 GCTATTAAAAAATACCTGTCTACAGATTGCCCTACCCCAAGTCAGTGTGTGGCCCCGA 1509
 Db |||||
 QY 1510 ACCTTAGGCAAAACAGCAAACTGTCTATGGCCATTGTCAAAAGATTGCCCTACAGATGAAC 1569
 Db |||||
 QY 1570 TGAAGATGGAGGAGAGCTCTGAGGGTGGACATCCCTGAACTCTGTGATGATGTT 1629
 Db |||||
 QY 1630 GGATCGATTGTTTACCATGACATGACAGCTGGCGGAGGTCAATCGCAGGATTGTTGCC 1689
 Db |||||
 QY 1690 AGCATCAATGAAGGATGACCCGCTGTTCTACGCTGCATATTTTCAGGATAGAGACAG 1749
 Db |||||
 QY 1750 GAGCTGTAGATGGGCTCAAGTCTCCTGCAAGCGGCTCTGAGGGCTTGGATAGCTGC 1809
 Db |||||
 QY 1810 AATGACTACATGCCAGCGGATCATCGTGTACCGGATGGCGTA 1854
 Db |||||
 QY 1877 AATGACTACATGCCAGCGGATCATCGTGTACCGGATGGCGTA 407

RESULT 6
 AA430311 587 bp mRNA linear EST 16-OCT-1997
 LOCUS ZW68H01.r1 Soares testis_NHT Homo sapiens cDNA clone IMAGE:781393
 DEFINITION 5' similar to WP:00105.2 CE06748 ; mRNA sequence.
 ACCESSION AA430311
 VERSION AA430311.1 GI:2113521
 KEYWORDS EST.
 SOURCE Homo sapiens (human)
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 REFERENCE 1 (bases 1 to 587)
 AUTHORS Hillier, L., Allen, M., Bowles, L., Dubucque, T., Geisel, G., Jost, S.,
 Kucaba, T., Lacy, M., Le, N., Lennon, G., Marra, M., Martin, J., Moore, B.,
 Schellenger, K., Steptoe, M., Tan, F., Theising, B., White, Y., Wylie,
 T., Waterston, R., and Wilson, R.
 JOURNAL WashU-Merck EST Project 1997
 COMMENT Contact: Wilson RK
 Washington University School of Medicine
 4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
 Tel: 314 286 1800
 Fax: 314 286 1810
 Email: est@wustl.edu
 This clone is available royalty-free through LLNL; contact the
 IMAGE Consortium (info@image.llnl.gov) for further information.
 Seq primer: -28m3 rev2 ET from AmerSham
 High quality sequence stop: 498.
 Location/Qualifiers
 1. .587
 /organism="Homo sapiens"
 /mol_type="mRNA"
 /db_xref="taxon:9606"
 /clone="IMAGE:781393"

```
/sex="male"
/lab_host="DH10B"
/clone_lib="Soares testis NHT"
/notes="Vector: pMT3D-Pac (Pharmacia) with a modified
polylinker; Site_1: Not I; Site_2: Eco RI; 1st strand cDNA
was prepared from mRNA obtained from Clontech Laboratories
, Inc., and primed with a Not I - oligo(dT) primer [5',
TGTTACCAATCTGAAGTGGAGCGCGCCCAATTTTGTTTT 3'].
Double-stranded cDNA was ligated to Eco RI adaptors
(Pharmacia), digested with Not I and cloned into the Not I
and Eco RI sites of the modified pMT3 vector. Library
went through one round of normalization to Cot5, and was
constructed by Bento Soares and M. Fatima Bonaldo. "
```

BASE COUNT 191 a 122 c 136 g 138 t

Query Match 15.5%; Score 361; DB 9; Length 587;
Best Local Similarity 100.0%; Pred. No. 2.7e-185;
Matches 361; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1852 GTAGGACGCGCCAGCTGAAACACTGGTGAACCTACGAGTGCACAGTTTGGATTGT 1911
Db |||||||
QY 38 GTAGGACGCGCCAGCTGAAACACTGGTGAACCTACGAGTGCACAGTTTGGATTGT 97
Db |||||||

QY 1912 CTAATCCATTTGTAGAGTTTCAACACCTAGACTAAGCGTAAATTTGGTGAAGAAAGA 1971
Db |||||||

QY 98 CTAATCCATTTGTAGAGTTTCAACACCTAGACTAAGCGTAAATTTGGTGAAGAAAGA 157
Db |||||||

QY 1972 GTGACACCAAGATTTTGTCTAGCTGTGGAGGAAGACTTCAGAAATCCATTCCTGGGAACA 2031
Db |||||||

QY 158 GTGACACCAAGATTTTGTCTAGCTGTGGAGGAAGACTTCAGAAATCCATTCCTGGGAACA 217
Db |||||||

QY 2032 GTTATTGATGTAGAGTTTACAGACCAAGATGGTATGACTTTTATTCGTGAGCCAGGCT 2091
Db |||||||

QY 218 GTTATTGATGTAGAGTTTACAGACCAAGATGGTATGACTTTTATTCGTGAGCCAGGCT 277
Db |||||||

QY 2092 GTGAGAAAGTGTAGTGTCTTCCACACATTAACAATGTCTATGATCAACACAGCGGCGCT 2151
Db |||||||

QY 278 GTGAGAAAGTGTAGTGTCTTCCACACATTAACAATGTCTATGATCAACACAGCGGCGCT 337
Db |||||||

QY 2152 AAGCCAGACACATACAGCGCTTGACTCAAGCTGTGCCACATCTATTACAACTGSCCA 2211
Db |||||||

QY 338 AAGCCAGACACATACAGCGCTTGACTCAAGCTGTGCCACATCTATTACAACTGSCCA 397
Db |||||||

2212 G 2212
398 G 398

RESULT 7
AA969938
LOCUS
DEFINITION
OP60a11.s1 Soares_NFL_T_GBC_S1 Homo sapiens cDNA clone
IMAGE:1581212.3 similar to WP:00165.2 CB06748 ;, mRNA sequence.
AA969938
ACCESSION
VERSION
KEYWORDS
SOURCE
ORGANISM
Homo sapiens (human)
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 462)
NCI-CCAP http://www.ncbi.nlm.nih.gov/ncicgap.
National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
Tumor Gene Index
Unpublished
Contact: Robert Strausberg, Ph.D.
Email: cgapbs@mail.nih.gov
This clone is available royalty-free through LLNL; contact the
IMAGE Consortium (info@image.llnl.gov) for further information.
Insert length: 521 Std Error: 0.00
Seq primer: -40m13 fwd. ET from Amersham.
Location/Qualifiers

```
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="IMAGE:1581212"
/lab_host="DH10B"
/clone_lib="Soares_NFL_T_GBC_S1"
/notes="Organ: pooled; Vector: pMT3D-Pac (Pharmacia) with
a modified polylinker; Site_1: Not I; Site_2: Eco RI;
Equal amounts of plasmid DNA from three normalized
libraries (fetal lung NbHL19W, testis NHT, and B-cell
NCI CCAP GC31) were mixed, and ss circles were made in
vitro. Following HAP purification, this DNA was used as
tracer in a subtractive hybridization reaction. The driver
was PCR-amplified cDNAs from pools of 5,000 clones made
from the same 3 libraries. The pools consisted of
I.M.A.G.E. clones 297480-302087, 682632-687239,
726408-728711, and 729096-731399. Subtraction by Bento
Soares and M. Fatima Bonaldo. "
```

BASE COUNT 127 a 97 c 124 g 114 t

ORIGIN

Query Match 13.8%; Score 321; DB 9; Length 462;
Best Local Similarity 99.8%; Pred. No. 1.9e-163;
Matches 441; Conservative 0; Mismatches 0; Indels 1; Gaps 1;

QY 1714 TGGTTCTCAGCTGCATATTTTCAGGATAGAGACAGAGCTGGTAGATGGCTCAAAGTC 1773
Db |||||||

QY 1 1 TGGTTCTCAGCTGCATATTTTCAGGATAGAGACAGAGCTGGTAGATGGCTCAAAGTC 60
Db |||||||

QY 1774 TGCCTCAAGCGGCTCTGAGGCTTGGAAATAGCTCAATGAGTACATGCCAGCCGATC 1833
Db |||||||

QY 61 TGCCTCAAGCGGCTCTGAGGCTTGGAAATAGCTCAATGAGTACATGCCAGCCGATC 120
Db |||||||

QY 1834 ATCGTGTACCGGATCGCTAGGACGCGCCAGCTCAAAACACTGTGTAACACGAGTG 1893
Db |||||||

QY 121 ATCGTGTACCGGATCGCTAGGACGCGCCAGCTCAAAACACTGTGTAACACGAGTG 179
Db |||||||

QY 1894 CCACAGTTTGGATTGTCTAAATCCATTTGGTAGAGTTACAACTAGACTAAACGGTA 1953
Db |||||||

QY 180 CCACAGTTTGGATTGTCTAAATCCATTTGGTAGAGTTACAACTAGACTAAACGGTA 239
Db |||||||

QY 1954 ATTGTGCTGAAGAAAGAGTGAACACAGATTTTTCAGTCTGAGGAGACATTCAG 2013
Db |||||||

QY 240 ATTGTGCTGAAGAAAGAGTGAACACAGATTTTTCAGTCTGAGGAGACATTCAG 259
Db |||||||

QY 2014 AATCCACTTCTCGGAACAGTTTATGATGTAGAGTTTACCAGACCAAGATGGTATGACTTT 2073
Db |||||||

QY 300 AATCCACTTCTCGGAACAGTTTATGATGTAGAGTTTACCAGACCAAGATGGTATGACTTT 359
Db |||||||

QY 2074 TTTATGCTGAGCCAGCTGTGAGAGTGTAGTGTTCCTCCACACATTAACATGTCATC 2133
Db |||||||

QY 360 TTTATGCTGAGCCAGCTGTGAGAGTGTAGTGTTCCTCCACACATTAACATGTCATC 419
Db |||||||

QY 2134 TATGACAAACAGCGGCTCGAAGC 2155
Db |||||||

QY 420 TATGACAAACAGCGGCTCGAAGC 441
Db |||||||

RESULT 8
BI560710
LOCUS
DEFINITION
60325474.F1 NIH_MGC_97 Homo sapiens cDNA clone IMAGE:5296954 5',
mRNA sequence.
BI560710
ACCESSION
VERSION
KEYWORDS
SOURCE
ORGANISM
Homo sapiens (human)
Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 779)
NIH-MGC http://mgc.ncbi.nih.gov/.

TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished
COMMENT Contact: Robert Strausberg, Ph.D.
 Email: cgapbs@mail.nih.gov
 Tissue Procurement: Miklos Palkovits, M.D., Ph.D.
 CDNA Library Preparation: Michael J. Brownstein (NHGRI), Shiraki Toshiyuki and Piero Carninci (RIKEN)
 CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
 DNA Sequencing by: Incyte Genomics, Inc.
 Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at: http://image.llnl.gov
 Plate: LLAM1751 row: h column: 11
 High quality sequence stop: 732.
FEATURES
 Location/Qualifiers
 1..779
 /organism="Homo sapiens"
 /mol_type="mRNA"
 /db_xref="taxon:9606"
 /clone="IMAGE:5296954"
 /lab_host="DH10B"
 /clone_lib="NIH MGC 97"
 /note="Organ: testis; Vector: pBluescriptR (modified pBluescript KS+); Site 1: BamHI; Site 2: SalI-XhoI (gtcgag); Oligo-dT primed using primer 5'-TTTTTTTTTTTTTTVN-3', size-selected for average insert size 2.2 kb and normalized to ROT 5. This is a primary library enriched for full-length clones and constructed using the Cap-trapper method (Carninci, in preparation). Library constructed by M. Brownstein (NIH/NHGRI, National Institutes of Health). Note: this is a NIH_MGC Library."
 234 a 159 c 205 g 181 t
 BASE COUNT
 ORIGIN
 Query Match 13.7%; Score 318; DB 12; Length 779;
 Best Local Similarity 100.0%; Pred. No. 9.7e-162;
 Matches 318; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 32 TAGACCATGTTAAGAAATCAAAACAGGTTCTTCAGGCATTATAGTAAGGTTAAGCACTA 91
 Db 368 TAGACCATGTTAAGAAATCAAAACAGGTTCTTCAGGCATTATAGTAAGGTTAAGCACTA 427
 QY 92 ACCATTTCGGCTGACATCCCGTCCCGCCAGTGGGCTTATATCAGTATCAGTATGACTATA 151
 Db 428 ACCATTTCGGCTGACATCCCGTCCCGCCAGTGGGCTTATATCAGTATGACTATA 487
 QY 152 ACCCACTGATGGAAGCCAGAAAGACTCCGTTTCAGTCTTCTTTTCAACGAGATCTAA 211
 Db 488 ACCCACTGATGGAAGCCAGAAAGACTCCGTTTCAGTCTTCTTTTCAACGAGATCTAA 547
 QY 212 TTGGAAGTGTATGCTTTTGTAGGAACGATATATTTTACCTTAAAGACTACAGCAA 271
 Db 548 TTGGAAGTGTATGCTTTTGTAGGAACGATATATTTTACCTTAAAGACTACAGCAA 607
 QY 272 AGTTACTGAAGTTTATAGTAAGCCGGAATGGAGGATGTGAGGATTAACGATCATT 331
 Db 608 AGTTACTGAAGTTTATAGTAAGCCGGAATGGAGGATGTGAGGATTAACGATCATT 667
 QY 332 TAACAAATGAATCTCCAC 349
 Db 668 TAACAAATGAATCTCCAC 685

RESULT 9
 BG773137
 LOCUS 60271862F1 NIH_MGC_97 Homo sapiens CDNA clone IMAGE:4838918 5',
 DEFINITION mRNA sequence.
 BG773137
 BG773137.1 GI:14083790
 VERSION EST.
 KEYWORDS Homo sapiens (human)
 SOURCE Homo sapiens
 ORGANISM Homo sapiens

REFERENCE 1 (bases 1 to 818)
AUTHORS NIH-MGC http://mgc.nci.nih.gov/
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished
COMMENT Contact: Robert Strausberg, Ph.D.
 Email: cgapbs@mail.nih.gov
 Tissue Procurement: Miklos Palkovits, M.D., Ph.D.
 CDNA Library Preparation: Michael J. Brownstein (NHGRI), Shiraki Toshiyuki and Piero Carninci (RIKEN)
 CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
 DNA Sequencing by: Incyte Genomics, Inc.
 Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at: http://image.llnl.gov
 Plate: LLAM10773 row: k column: 15
 High quality sequence stop: 789.
FEATURES
 Location/Qualifiers
 1..818
 /organism="Homo sapiens"
 /mol_type="mRNA"
 /db_xref="taxon:9606"
 /clone="IMAGE:4838918"
 /lab_host="DH10B"
 /clone_lib="NIH MGC 97"
 /note="Organ: testis; Vector: pBluescriptR (modified pBluescript KS+); Site 1: BamHI; Site 2: SalI-XhoI (gtcgag); Oligo-dT primed using primer 5'-TTTTTTTTTTTTTTVN-3', size-selected for average insert size 2.2 kb and normalized to ROT 5. This is a primary library enriched for full-length clones and constructed using the Cap-trapper method (Carninci, in preparation). Library constructed by M. Brownstein (NIH/NHGRI, National Institutes of Health). Note: this is a NIH_MGC Library."
 245 a 174 c 198 g 201 t
 BASE COUNT
 ORIGIN
 Query Match 13.2%; Score 308; DB 12; Length 818;
 Best Local Similarity 100.0%; Pred. No. 2.9e-156;
 Matches 308; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 32 TAGACCATGTTAAGAAATCAAAACAGGTTCTTCAGGCATTATAGTAAGGTTAAGCACTA 91
 Db 329 TAGACCATGTTAAGAAATCAAAACAGGTTCTTCAGGCATTATAGTAAGGTTAAGCACTA 388
 QY 92 ACCATTTCGGCTGACATCCCGTCCCGCCAGTGGGCTTATATCAGTATCAGTATGACTATA 151
 Db 389 ACCATTTCGGCTGACATCCCGTCCCGCCAGTGGGCTTATATCAGTATCAGTATGACTATA 448
 QY 152 ACCCACTGATGGAAGCCAGAAAGACTCCGTTTCAGTCTTCTTTTCAACGAGATCTAA 211
 Db 449 ACCCACTGATGGAAGCCAGAAAGACTCCGTTTCAGTCTTCTTTTCAACGAGATCTAA 508
 QY 212 TTGGAAGTGTATGCTTTTGTAGGAACGATATATTTTACCTTAAAGACTACAGCAA 271
 Db 509 TTGGAAGTGTATGCTTTTGTAGGAACGATATATTTTACCTTAAAGACTACAGCAA 568
 QY 272 AGTTACTGAAGTTTATAGTAAGCCGGAATGGAGGATGTGAGGATTAACGATCATT 331
 Db 569 AGTTACTGAAGTTTATAGTAAGCCGGAATGGAGGATGTGAGGATTAACGATCATT 628
 QY 332 TAACAAAT 339
 Db 629 TAACAAAT 636

RESULT 10
 AL705072
 LOCUS DKFP686E0634 r1 686 (synonym: hlec3) Homo sapiens cDNA clone
 DEFINITION DKFP686E0634 5', mRNA sequence.
 AL705072
 AL705072 622 bp mRNA linear EST 22-MAR-2002

```

VERSION AL705072.1 GI:19688427
SOURCE EST.
ORGANISM Homo sapiens (human)
REFERENCE Bukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
AUTHORS Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
1 (bases 1 to 622)
TITLE Otenwaelder,B., Obermaier,B., Mewes,W., Mewes,H.W., Weil,B. and Wiemann
S.)
JOURNAL Unpublished
COMMENT Contact: Otenwaelder B
MIPS Ingolstaedter Landstr.1, D-85764 Neuherberg, Germany
This is the 5' sequence of the clone insert
Clone from S. Wiemann, Molecular Genome Analysis, German Cancer
Research Center (DKFZ); Email s.wiemann@dkfz-heidelberg.de;
sequenced by MediGenomix (Martinsried/Germany) within the cDNA
sequencing consortium of the German Genome Project. No sl sequence
available.
This clone (DKFZp686E0634) is available at the RZPD in Berlin.
Please contact the RZPD: Ressourcenzentrum, Heubnerweg 6, 14059
Berlin- Charlottenburg, GERMANY; Email: clone@rzpd.de.
FEATURES
Location/Qualifiers
1..622
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="DKFZp686E0634"
/tissue_type="human skeletal muscle"
/dev_stage="adult"
/lab_host="DH10B"
/clone_lib="686 (synonym: blcc3)"
/note="Vector: pTriplex2; Site_1: SfiIA, Site_2: SfiIB;
cDNA-collection"
BASE COUNT 187 a 134 c 168 g 133 t
ORIGIN
Query Match 12.2%; Score 285; DB 9; Length 622;
Best Local Similarity 100.0%; Pred. No. 16-143;
Matches 285; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 32 TAGACCATGTTAAGGAATCAAAACAGGTTCTTCAGGCATTATAGTAAGTTAAGCACTA 91
Db 338 TAGACCATGTTAAGGAATCAAAACAGGTTCTTCAGGCATTATAGTAAGTTAAGCACTA 397
QY 92 ACCATTTCGGCTGACATCCCGTCCCGAGTGGCCCTTATATCATGATCATTCATGACTATA 151
Db 398 ACCATTTCGGCTGACATCCCGTCCCGAGTGGCCCTTATATCATGATCATTCATGACTATA 457
QY 152 ACCCATCTGATGGAAGCAGAGACTCCGTTTCAGCTCTCTCTTTTCAACGAGAGTCTAA 211
Db 458 ACCCATCTGATGGAAGCAGAGACTCCGTTTCAGCTCTCTCTTTTCAACGAGAGTCTAA 517
QY 212 TTGGAAGTGTCTATGCTTTTGATGGAACGATATTTTACCTAAAGACTACACAAA 271
Db 518 TTGGAAGTGTCTATGCTTTTGATGGAACGATATTTTACCTAAAGACTACACAAA 577
QY 272 AGTTACTGAAGTTTATGTAAGACCCGGAATGGAGAGATGGA 316
Db 578 AGTTACTGAAGTTTATGTAAGACCCGGAATGGAGAGATGGA 622
RESULT 11
BM728018
LOCUS 559 bp mRNA linear EST 01-MAR-2002
DEFINITION UI-E-EJO-ait-h-19-0-UI.r1 UI-E-EJO Homo sapiens cDNA clone
BM728018
ACCESSION UI-E-EJO-ait-h-19-0-UI 5', mRNA sequence.
VERSION BM728018.1 GI:19049351
KEYWORDS EST.
SOURCE Homo sapiens (human)

```

```

ORGANISM Homo sapiens
Bukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
1 (bases 1 to 559)
REFERENCE Bonaldo,M.F., Lennon,G. and Soares,M.B.
AUTHORS Normalization and subtraction: two approaches to facilitate gene
discovery
JOURNAL Genome Res. 6 (9), 791-806 (1996)
MEDLINE 97044477
PUBMED 889548
COMMENT Contact: Soares, MB
Coordinated Laboratory for Computational Genomics
University of Iowa
375 Newton Road, 4156 MEBRF, Iowa City, IA 52242, USA
Tel: 319 335 8250
Fax: 319 335 9565
Email: bento-soares@uiowa.edu
Tissue Procurement: Dr. Gregg Hageman
cDNA Library preparation: Dr. M. Bento Soares, University of Iowa
cDNA Library Arrayed by: Dr. M. Bento Soares, University of Iowa
DNA Sequencing by: Dr. M. Bento Soares, University of Iowa
Clone Distribution: Researchers may obtain clones from Research
Genetics (www.resgen.com).
Seq primer: M13 Reverse.
FEATURES
Location/Qualifiers
1..559
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="UI-E-EJO-ait-h-19-0-UI"
/tissue_type="fetal eyes, lens, eye anterior segment,
optic nerve, retina, Retina Foveal and Macular, RPE and
Choroid"
/dev_stage="fetal and adult"
/lab_host="DH10B (Life Technologies) (T1 phage resistant)"
/clone_lib="UI-E-EJO"
/note="Organ: eye; Vector: pT73-Pac (Pharmacia) with a
modified polylinker; Site_1: EcoR I; Site_2: Not I;
UI-E-EJO is a subtracted cDNA library constructed
according to Bonaldo, Lennon and Soares, Genome Research,
6:791-806, 1996. First strand cDNA synthesis was primed
with an oligo-dT primer containing a Not I site. Double
stranded cDNA was ligated to an EcoR I adaptor, digested
with Not I, and cloned directionally into pT73-Pac
vector. The oligonucleotide used to prime the synthesis of
first-strand cDNA contains a library tag sequence that is
located between the Not I site and the (dT)18 tail. The
sequence tags for this library are: fetal eyes, AGAATCAAGA
; lens, CGATTAGCGA; eye anterior segment, AATGCCCAT;
optic nerve, CCATTAGTG; retina, CCGCG; Retina Foveal and
Macular, GTCC; RPE and Choroid, ACCTA. This library was
created for the program, Gene Discovery in the Visual
System, supported by National Eye Institute (NEI)."
BASE COUNT 170 a 114 c 144 g 130 t
ORIGIN
Query Match 11.8%; Score 275; DB 12; Length 559;
Best Local Similarity 100.0%; Pred. No. 2.9e-138;
Matches 275; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1938 CCTAGACTAAGCGTAATTGTGGTGAAGAAAGAGTGAACACAGATTTTGTCTCAGTC 1997
Db 212 CCTAGACTAAGCGTAATTGTGGTGAAGAAAGAGTGAACACAGATTTTGTCTCAGTC 271
QY 1998 TGGAGGAAGACTTCAGATTCACATTCCTTGGACACTTATGATGATGAGTTACAGACC 2057
Db 272 TGGAGGAAGACTTCAGATTCACATTCCTTGGACACTTATGATGATGAGTTACAGACC 331
QY 2058 AGAATGGTATGACTTTTATCGTGAGCCAGCTCTGTGAGAGTGTGTATGTTCTCCAC 2117
Db 332 AGAATGGTATGACTTTTATCGTGAGCCAGCTCTGTGAGAGTGTGTATGTTCTCCAC 391
QY 2118 ACATTACAATGTCTCTATGACACAGCGCTTGAAGCCAGCCACATACAGCGCTTGAC 2177

```

```
Db 392 ACATTCAATGTTCATCTATGACAAACGCGCCCTGAAGCCACACCATACATACAGCGCTTGAC 451
QY 2178 CTACAAAGCTGTGCCACATCTATTACAACTGGCCAG 2212
Db 452 CTACAAAGCTGTGCCACATCTATTACAACTGGCCAG 486

RESULT 12
BQ374910 420 bp mRNA linear EST 21-MAY-2002
LOCUS BQ374910
DEFINITION MR4-TN0109-060900-101-h03 TN0109 Homo sapiens cDNA, mRNA sequence.
ACCESSION BQ374910
VERSION BQ374910.1 GI:21050424
KEYWORDS EST.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
AUTHORS Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 420)
Dias Neto,E., Garcia Correa,R., Verjovski-Almeida,S., Briones,M.R.,
Nagai,M.A., da Silva,W. Jr., Zago,M.A., Bordin,S., Costa,F.F.,
Goldman,G.H., Carvalho,A.P., Matsukuma,A., Baia,G.S., Simpson,D.H.,
Brunstein,A., deOliveira,P.S., Bucher,P., Jongeneel,C.V., O'Hare
M.J., Soares,F., Brentani,R.R., Reis,L.F., de Souza,S.J. and
Simpson,A.J.
TITLE Shotgun sequencing of the human transcriptome with ORF expressed
sequence tags
JOURNAL Proc. Natl. Acad. Sci. U.S.A. 97 (7), 3491-3496 (2000)
MEDLINE 20202663
PUBMED 10737800
COMMENT Contact: Simpson A.J.G.
Laboratory of Cancer Genetics
Ludwig Institute for Cancer Research
Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP,
Brazil
Tel: +55-11-2704922
Fax: +55-11-2707001
Email: asimpson@ludwig.org.br
This sequence was derived from the FAPESP/LICR Human Cancer Genome
Project. This entry can be seen in the following URL
(http://www.ludwig.org.br/scripts/gethtml2.pl?tl=MR4&tl2=MR4-TN0109-
060900-101-h03&tl3=2000-09-08&tl4=1)
Seq primer: puc 18 forward
High quality sequence stop: 420.
FEATURES
Location/Qualifiers
1..420
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/dev_stage="Adult"
/clone_lib="TN0109"
/note="Organ: testis normal; Vector: puc18; Site 1: SmaI;
Site 2: SmaI; A mini-library was made by cloning products
derived from ORFESTES PCR (U.S. Letters Patent application
No.196,716 - Ludwig Institute for Cancer Research)
profiles into the puc 18 vector. Reverse transcription of
tissue mRNA and cDNA amplification were performed under
low stringency conditions."
BASE COUNT 122 a 94 c 100 g 104 t
ORIGIN
Query Match 11.0%; Score 257; DB 13; Length 420;
Best Local Similarity 99.7%; Pred. No. 1.9e-128;
Matches 307; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1905 GGATTGCTCTAAATCCATTCGTAGAGTTACACCCCTAGACTACCGTAATGTTGGTGA 1964
Db 97 GGATTGCTCTAAATCCATTCGTAGAGTTACACCCCTAGACTACCGTAATGTTGGTGA 156

QY 1965 GAAAGAGTGAACACACAGATTTTTCCTAGTCTGGAGGAGACTTCAGATCCATCTCC 2024
Db 157 SAAAGAGTGAACACACAGATTTTTCCTAGTCTGGAGGAGACTTCAGATCCATCTCC 216
```

```
QY 2025 TCGAACAGTATTGATGTAGAGTTTACAGACCAAGATGGTATGACTTTTATCTGTGAG 2084
Db 217 TCGAACAGTATTGATGTAGAGTTTACAGACCAAGATGGTATGACTTTTATCTGTGAG 276
QY 2085 CCAGGCTGTGAGAAGTGGTGTGTTTCTCCACACATTACATGTCATCTATGACACAG 2144
Db 277 CCAGGCTGTGAGAAGTGGTGTGTTTCTCCACACATTACATGTCATCTATGACACAG 336
QY 2145 CGGCTCTGAAGCCAGACCATACATACAGCGCTTGACCTACAGCTGTGCCACATCTATTACAA 2204
Db 337 CGGCTCTGAAGCCAGACCATACATACAGCGCTTGACCTACAGCTGTGCCACATCTATTACAA 396
QY 2205 CTGGCCAG 2212
Db 397 CTGGCCAG 404

RESULT 13.
BQ435882 718 bp mRNA linear EST 24-MAY-2002
LOCUS BQ435882
DEFINITION AGENCOURT 7897283 NIH_MGC_82 Homo sapiens cDNA clone IMAGE:6103498
5', mRNA sequence.
ACCESSION BQ435882
VERSION BQ435882.1 GI:21174947
KEYWORDS EST.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
AUTHORS Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 718)
Nagai,M.A., da Silva,W. Jr., Zago,M.A., Bordin,S., Costa,F.F.,
Goldman,G.H., Carvalho,A.P., Matsukuma,A., Baia,G.S., Simpson,D.H.,
Brunstein,A., deOliveira,P.S., Bucher,P., Jongeneel,C.V., O'Hare
M.J., Soares,F., Brentani,R.R., Reis,L.F., de Souza,S.J. and
Simpson,A.J.
TITLE NIH-MGC http://mgc.ncl.nih.gov/.
JOURNAL National Institutes of Health, Mammalian Gene Collection (MGC)
COMMENT Unpublished
Contact: Robert Strausberg, Ph.D.
Email: cgapbs@mail.nih.gov
Tissue Procurement: CLONTECH
cDNA Library Preparation: CLONTECH Laboratories, Inc.
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Agencourt Bioscience Corporation
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
Plate: LLCM2339 row: n column: 11
High quality sequence stop: 518.
FEATURES
Location/Qualifiers
1..718
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="IMAGE:6103498"
/lab_host="DH10B (T1 phage-resistant)"
/clone_lib="NIH MGC 82"
/note="Organ: testis; Vector: pDNR-LIB (Clontech); Site 1:
SfiI (ggccatcgcc); Site 2: SfiI (ggccatcgcc); 5' and
3' adaptors were used in cloning as follows: 5' adaptor
sequence: 5'-CACGCCATATGCC-3' and 3' adaptor sequence:
5'-ATTCTAGAGCGCGCGGACATG-dt(30)BN-3' (where B = A,
C, or G and N = A, C, G, or T). Average insert size
1.35 kb (range 0.9-4.0 kb). 14/15 colonies contained
inserts by PCR. This library was enriched for full-length
clones and was constructed by Clontech Laboratories (Palo
Alto, CA)."
BASE COUNT 192 a 137 c 140 g 249 t
ORIGIN
Query Match 10.7%; Score 249; DB 13; Length 718;
Best Local Similarity 100.0%; Pred. No. 5.3e-124;
Matches 249; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2080 GTGACCCAGGCTGTGAGAAGTGGTATGTTTCTCCACACATTACATGTCATCTATGAC 2139
Db 5 GTGACCCAGGCTGTGAGAAGTGGTATGTTTCTCCACACATTACATGTCATCTATGAC 64
```

QY 2140 AACAGGGCTGAAGCCAGACACATACAGCGCTTGACCTACAGCTGTGCCACATCTAT 2199
Db 65 AACAGGGCTGAAGCCAGACACATACAGCGCTTGACCTACAGCTGTGCCACATCTAT 124
QY 2200 TAAACTGCGCCAGGTGTCATTCTGTCTCTGCTCCCTTGCCAGTACGCCACAGCTGGCT 2259
Db 125 TAAACTGCGCCAGGTGTCATTCTGTCTCTGCTCCCTTGCCAGTACGCCACAGCTGGCT 184
QY 2260 TTCTCTGTTGGCCAGAGTATTACAGAGAGCCAAATCTGTCACTGTCAAACGCCCTTTAC 2319
Db 185 TTCTCTGTTGGCCAGAGTATTACAGAGAGCCAAATCTGTCACTGTCAAACGCCCTTTAC 244
QY 2320 TACCTCTAA 2328
Db 245 TACCTCTAA 253

RESULT 14
BM680592/c
LOCUS
DEFINITION
ACCESSION
VERSION
KEYWORDS
SOURCE
ORGANISM
REFERENCE
AUTHORS
TITLE
JOURNAL
MEDLINE
PUBMED
COMMENT

BM680592 512 bp mRNA linear EST 27-FEB-2002
UI-B-EJ0-ait-h-19-0-UI.s1 UI-E-EJ0 Homo sapiens cDNA clone
UI-B-EJ0-ait-h-19-0-UI.3', mRNA sequence.
BM680592
BM680592.1 GI:18990488
EST.
Homo sapiens (human)
Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 512)
Bonaldo, M.F., Lennon, G. and Soares, M.B.
Normalization and subtraction: two approaches to facilitate gene
discovery
Genome Res. 6 (9), 791-806 (1996)
97044477
8899548
Contact: Soares, MB
Coordinated Laboratory for Computational Genomics
University of Iowa
375 Newton Road, 4156 MEBRF, Iowa City, IA 52242, USA
Tel: 319 335 8250
Fax: 319 335 9565
Email: bento-soares@uiowa.edu
Tissue procurement: Dr. Gregg Hageman
cDNA library prepared by: Dr. M. Bento Soares, University of Iowa
DNA Sequencing by: Dr. M. Bento Soares, University of Iowa
Clone Distribution: Researchers may obtain clones from Research
Genetics (www.resgen.com).
Seq primer: M13 Forward
POLYA=Yes.

FEATURES
source
1..512
/dev_stage="fetal and adult"
/lab_host="PH108 (Life Technologies) (T1 phage resistant)"
/clone_lib="UI-E-EJ0"
/note="Organ: eye; Vector: p7T73-Pac (Pharmacia) with a
modified polylinker; Site 1: EcoR I; Site 2: Not I;
UI-E-EJ0 is a subtracted cDNA library constructed
according to Bonaldo, Lennon and Soares, Genome Research,
6:791-806, 1996. First strand cDNA synthesis was primed
with an oligo-dT primer containing a Not I site. Double
stranded cDNA was ligated to an EcoR I adaptor, digested
with Not I, and cloned directionally into p7T73-Pac
vector. The oligonucleotide used to prime the synthesis of

first-strand cDNA contains a library tag sequence that is
located between the Not I site and the (dT)18 tail. The
sequence tags for this library are: fetal eyes, AGAATCAAGA
; lens, CGATTAGCGA; eye anterior segment, AATGCCGCAT;
optic nerve, CCATTAGATG; retina, CGCGG; Retina Foveal and
Macular, GTCC; RPE and Choroid, ACCTA. This library was
created for the program, Gene Discovery in the Visual
System, supported by National Eye Institute (NEI).

TAG LIB=UI-E-EJ0

TAG TISSUE=Foveal and Macular Retina

TAG_SEQ=GTCC

BASE COUNT 125 a 128 c 100 g 157 t 2 others
ORIGIN

Query Match 10.5%; Score 245; DB 12; Length 512;
Best Local Similarity 100.0%; Pred. No. 7.3e-122;
Matches 245; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1938 CCTAGACTTAACGGTAATTGTGTGAAGAAAGAGTGAACACAGATTTTGTCTCAGTC 1997
Db 344 CCTAGACTTAACGGTAATTGTGTGAAGAAAGAGTGAACACAGATTTTGTCTCAGTC 285
QY 1998 TGGAGGAAGACTTCAGAAATCCACTTCTTGGACAGCTTATTGATGTAGAGTTTACCAGACC 2057
Db 284 TGGAGGAAGACTTCAGAAATCCACTTCTTGGACAGCTTATTGATGTAGAGTTTACCAGACC 225
QY 2058 AGAATGGTATGACTTTTATCGTGAGCCAGGCTGTGAGAGTGTGTAGTGTTCCTCCAC 2117
Db 224 AGAATGGTATGACTTTTATCGTGAGCCAGGCTGTGAGAGTGTGTAGTGTTCCTCCAC 165
QY 2118 ACATTACAAATGTCATCTATGACACAGCGGCTGGAAGCCAGACACACATACACGGCTTGAC 2177
Db 164 ACATTACAAATGTCATCTATGACACAGCGGCTGGAAGCCAGACACATACACGGCTTGAC 105
QY 2178 CTACA 2182
Db 104 CTACA 100

RESULT 15
AL705301
LOCUS
DEFINITION
ACCESSION
VERSION
KEYWORDS
SOURCE
ORGANISM
REFERENCE
AUTHORS
TITLE
JOURNAL
COMMENT

AL705301 640 bp mRNA linear EST 22-MAR-2002
DKFZp686G1235 r1.686 (synonym: hicc3) Homo sapiens cDNA clone
DKFZp686G1235.5', mRNA sequence.
AL705301.1 GI:19688656
EST.
Homo sapiens (human)
Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 640)
Ottewaelder, B., Obermaier, B., Mewes, H.W., Weil, B. and
Wiemann, S.
EST (Ottewaelder, B., Obermaier, B., Mewes, H.W., Weil, B. and
Wiemann, S.)
Unpublished
Contact: Ottewaelder B
MIPS

Ingolstaedter Landstr.1, D-85764 Neuherberg, Germany
This is the 5' sequence of the clone insert
Clone from S. Wiemann, Molecular Genome Analysis, German Cancer
Research Center (DKFZ); Email s.wiemann@dkfz-heidelberg.de;
sequencing consortium of the German Genome Project. No s1 sequence
available.
This clone (DKFZp686G1235) is available at the RZPD in Berlin.
Please contact the RZPD: Ressourcencentrum, Heubnerweg 6, 14059
Berlin-Charlottenburg, GERMANY; Email: clone@rzpd.de.
Location/Qualifiers
1..640
/organism="Homo sapiens"
/mol_type="mRNA"

```
/db_xref="taxon:9606"  
/clone="DKFZp686G1235"  
/tissue_type="human skeletal muscle"  
/dev_stage="adult"  
/lab_host="DH10B"  
/clone_lib="686 (synonym: hlcc3)"  
/note="Vector: pTriplex2; Site_1: SfiIA; Site_2: SfiIB;  
cDNA-collection"  
BASE COUNT 190 a 139 c 174 g 137 t  
ORIGIN
```

```
Query Match 10.4%; Score 241; DB 9; Length 640;  
Best Local Similarity 100.0%; Pred.No. 1.2e-119;  
Matches 241; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
  
QY 32 TAGACCATGTTAAAGAAATCAAAAACAGGTTCTTCAGGCATTATAGTAAAGTTAAGCACTA 91  
|||  
Db 357 TAGACCATGTTAAAGAAATCAAAAACAGGTTCTTCAGGCATTATAGTAAAGTTAAGCACTA 416  
|||  
  
QY 92 ACCATTTCGGGCTGACATCCCGTCCCGAGTGGGCCCTTATACAGTATCATCATTGACTATA 151  
|||  
Db 417 ACCATTTCGGGCTGACATCCCGTCCCGAGTGGGCCCTTATACAGTATCATCATTGACTATA 476  
|||  
  
QY 152 ACCACTGTATGGAAGCCAGAAGACTCCGTTCCAGCTCTCTTTTCAACACGAAGATCTTAA 211  
|||  
Db 477 ACCACTGTATGGAAGCCAGAAGACTCCGTTCCAGCTCTCTTTTCAACACGAAGATCTTAA 536  
|||  
  
QY 212 TTGGAAGTGTATGCTTTTGTATGGAACGATATATTTTACCTAAAGACTACAGCAAA 271  
|||  
Db 537 TTGGAAGTGTATGCTTTTGTATGGAACGATATATTTTACCTAAAGACTACAGCAAA 596  
|||  
  
QY 272 A 272  
|  
Db 597 A 597
```

Search completed: December 6, 2003, 22:05:23
Job time : 4859 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: December 6, 2003, 09:02:07 ; Search time 8509 seconds
(without alignments)
11192.574 Million cell updates/sec

Title: US-10-043-774B-1

Perfect score: 2328

Sequence: 1 atgatctttgggtgaacac.....accgcctttactactcttaa 2328

Scoring table: IDENTITY NUC

Gapop 10.0 , Gapext 1.0

Searched: 2888711 segs, 20454813386 residues

Total number of hits satisfying chosen parameters: 5777422

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

GenEmbl :

1: gb_ba.*
2: gb_htg.*
3: gb_in.*
4: gb_om.*
5: gb_ov.*
6: gb_pat.*
7: gb_ph.*
8: gb_pl.*
9: gb_pr.*
10: gb_ro.*
11: gb_sts.*
12: gb_sy.*
13: gb_un.*
14: gb_vi.*
15: em_ba.*
16: em_fun.*
17: em_hum.*
18: em_in.*
19: em_mu.*
20: em_om.*
21: em_or.*
22: em_ov.*
23: em_pat.*
24: em_ph.*
25: em_pl.*
26: em_ro.*
27: em_sts.*
28: em_un.*
29: em_vi.*
30: em_htg_hum.*
31: em_htg_inv.*
32: em_htg_other.*
33: em_htg_mus.*
34: em_htg_pln.*
35: em_htg_rod.*
36: em_htg_mam.*
37: em_htg_vrt.*
38: em_sy.*
39: em_htgo_hum.*
40: em_htgo_mus.*
41: em_htgo_other.*

Pred. No. is the number of results predicted by chance to have a

score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	DB	ID	Description
1	2328	100.0	2328	9	AF264004	AF264004 Homo sapi
2	2316.2	99.5	3421	9	AF387507	AF387507 Homo sapi
3	2313	99.4	3362	9	AF104260	AF104260 Homo sapi
4	2313	99.4	3399	9	BC028581	BC028581 Homo sapi
5	2200.2	94.5	2579	9	AK093133	AK093133 Homo sapi
6	1684.6	72.4	3910	10	AB032604	AB032604 Mus muscu
7	1684.6	72.4	4064	10	AF438405	AF438405 Mus muscu
8	1013.2	43.5	3273	5	AF336369	AF336369 Danio rer
9	880	29.2	3006	9	BC031060	BC031060 Homo sapi
10	680	29.2	3138	9	AK093475	AK093475 Homo sapi
11	616.8	26.5	2456	9	AK096837	AK096837 Homo sapi
12	568.6	24.4	3057	3	AY014899	AY014899 Strongylo
13	503.2	21.6	3458	3	AK116906	AK116906 Ciona int
14	499.4	21.5	2605	3	AY014900	AY014900 Strongylo
15	460.6	19.8	4913	10	AB032605	AB032605 Mus muscu
16	439.8	18.9	3610	9	BC025995	BC025995 Homo sapi
17	435.2	18.7	3620	6	AX714218	AX714218 Sequence
18	435.2	18.7	3620	9	AK056418	AK056418 Homo sapi
19	381	16.4	3649	6	AX247684	AX247684 Sequence
20	381	16.4	3659	10	AF285586	AF285586 Mus muscu
21	379.4	16.3	1723	3	AY014901	AY014901 Strongylo
22	330.2	14.2	1455	5	BC049495	BC049495 Danio rer
23	330	14.2	2272	6	BD157952	BD157952 Primer fo
24	330	14.2	2272	9	AK027497	AK027497 Homo sapi
25	310.6	13.3	2761	3	AF145680	AF145680 Drosophil
26	257	11.0	3047	3	AF104354	AF104354 Drosophil
27	237.2	10.2	1733	6	BD156024	BD156024 Primer fo
28	222.4	10.2	1733	9	AK001213	AK001213 Homo sapi
29	222.4	9.6	63578	9	AC127071	AC127071 Homo sapi
30	220.8	9.5	193697	2	AC025837	AC025837 Homo sapi
31	203.4	8.7	61995	2	AC090565	AC090565 Homo sapi
32	184	7.9	7781	3	DM0NARPL9	X94613 D.melanogas
33	184	7.9	37980	3	AC005116	AC005116 Drosophil
34	184	7.9	78507	2	AC019704	AC019704 Drosophil
35	184	7.9	169210	3	AC092226	AC092226 Drosophil
36	184	7.9	176306	3	AC007186	AC007186 Drosophil
37	184	7.9	289460	3	AE003630	AE003630 Drosophil
38	182.4	7.8	901	3	AF334408	AF334408 Drosophil
39	171.8	7.4	548	6	AF387649	AF387649 Sequence
40	170.6	7.3	186252	2	AC116715	AC116715 Mus muscu
41	170.6	7.3	211319	2	AC111089	AC111089 Mus muscu
42	164.4	7.1	259213	2	AC096050	AC096050 Rattus no
43	161.4	6.9	3677	2	AC017885	AC017885 Drosophil
44	142.2	6.1	9359	3	AF104355	AF104355 Drosophil
45	138.4	5.9	39365	2	AC100141	AC100141 Mus muscu

ALIGNMENTS

RESULT 1
AF264004
LOCUS
DEFINITION
ACCESSION
VERSION
KEYWORDS
SOURCE
ORGANISM
REFERENCE
AUTHORS
TITLE

AF264004
Homo sapiens
AF264004
AF264004.1
GI:15216446
Homo sapiens (human)
Homo sapiens
Mammalia; Eutheria;
1 (bases 1 to 2328)
Sharma,A.K., Nelson,M.C., Brandt,J.E., Wessman,M., Mahmud,N.,
Weller,K.P. and Hoffman,R.
Human CD34+ stem cells express the hiwi gene, a human homologue of

2328 bp
mRNA
linear
PRI 30-APR-2003
complete cds.

[illegible]

RESULT 2
AF387507 LOCUS
AF387507 DEFINITION
AF387507 Homo sapiens PIWI protein mRNA, complete cds.
AF387507 VERSION
AF387507.1 GI:14579644
AF387507 SOURCE
Homo sapiens (human)
Homo sapiens ORGANISM
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 3421) REFERENCE
Sha, J.H. 1
Cloning and identification of human piwi protein related to testis development TITLE
2 (bases 1 to 3421) JOURNAL
Sha, J.H., Li, J.M. and Zhou, Z.M. Unpublished
Submitted (30-MAY-2001) Key Lab of Reproductive Medicine, Nanjing Medical University, 140 Han Zhong Road, Nanjing Medical University, China
Direct Submission JOURNAL

Jiangsu 210029, China
 Location/Qualifiers
 1. .3421
 /organism="Homo sapiens"
 /mol_type="mRNA"
 /db_xref="taxon:9606"
 /tissue_type="testis"
 /dev_stage="adult"
 64. .2649
 /note="higher expression in adult testis than 6 month old
 embryo testis"
 /codon_start=1
 /product="PIWI protein"
 /protein_id="AAK69348.1"
 /db_xref="GI:14579645"
 /translation="MTGARARARGARGOETALVGVSTASQOQPIQPPQPAPPA
 ELFGRGORGTAGTAKSOGLOISAGFQLSLAERGRERDFHDLGVNTRQLNDHVK
 SKTSGSIIVRSLTNHFRILSPQWALYQHIDYNPLMEARLRLSALLFOHEDLIGK
 HAFDTILFLPKLQKQIVTEVSKTRNGEDVRITLITNELPSTPCLQFYNIIFRR
 LUKIMNLQIGRNYNPNDPIIDPSHRLVWPGFTTSILQYENSIMLCTDQNPKSTFK
 ETVLDMFNFYHQTKEHQEQVSKELIGLVLTIKNNKTRVDDIDWDQNPKSTFK
 ADGSEVFLFYRKQVQNIITDLKQPVLSQPKRRPGGTLPGPAMLIPELCYLTG
 TDKMRNDFNMKDLAVHTSLTPEQOREVGRGLDYTHKNDVORELFDWGLSDPSNLI
 SSGRILQTEKHQGGTKDYNDPQPADNSKETRGAPLISVYKPLDNWLLIITRRN
 NSQLNLTQIPANGQMRKAMIEVDRTAEYLRVQQQVADTQILVCLLSNRK
 KYDAIKKLYCTDPTSCQCVATLTKGQCTWMAITKIALQNMKGGELEWVDDIPLK
 LVYIVGIDCVDMTAGRRGIAGTAVASINEGMTWFSRCIFQDRGEQLVDGLKVLQAA
 LRAWNSQYMPRSIIIVYRDGVGDGLKTLVNYEVQFLDCLKSIIGRGNPRLTIVV
 KKEVNTPEAQSGRGLNPPLPGTVIDVEVTRPEWYDFFIVSQAVRGSVSPTHNYVY
 DNSGLKPDHIQRUTYLKCHIIYNWPGVIRVPAPQYAHKLAFLVQSGSIHREPNLSLS
 RLYYL"
 polyA_signal 3354. .3359
 BASE_COUNT 1071 a 661 c 767 g 922 t
 ORIGIN

 Query Match 99.5%; Score 2316.2; DB 9; Length 3421;
 Best Local Similarity 99.9%; Pred. No. 0;
 Matches 2318; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

 QY 8 TTGGTGTGAACAAGGCGAAGCTTAGACCATGTTAAAGAAATCAAAACAGGTTCTTCAG 67
 Db 329 TTGGTGTGAATAAAGGCGAGACTTAGACCATGTTAAAGAAATCAAAACAGGTTCTTCAG 388
 QY 68 GCATTATAGTAAGGTTAAGCACTAAACCATTTCCGGCTGACATCCGCTCCCGAGTGGCGCT 127
 Db 389 GCATTATAGTAAGGTTAAGCACTAAACCATTTCCGGCTGACATCCGCTCCCGAGTGGCGCT 448
 QY 128 TATATCATGATCATCATTGACTATAAACCACGTGATGGAAGCCAGAGACTCCGTTCACTC 187
 Db 449 TATATCAGTATCATCATTGACTATAAACCACGTGATGGAAGCCAGAGACTCCGTTCACTC 508
 QY 188 TTCTTTTTCACACGAGAGATCTAATTGGAAAGTGTCAATCTTTTGTATGGAACAGTATTAT 247
 Db 509 TTCTTTTTCACACGAGAGATCTAATTGGAAAGTGTCAATCTTTTGTATGGAACAGTATTAT 568
 QY 248 TTTTACCTAAAAGACTACAGCAAAAGGTTTACTGAAGTTTTTAGTAAAGCCGGAATGAG 307
 Db 569 TTTTACCTAAAAGACTACAGCAAAAGGTTTACTGAAGTTTTTAGTAAAGCCGGAATGAG 628
 QY 308 AGGATGTGAGGATAACGATCACTTTAAACAAATGAACTTCCACCTACATCACCACCTTGT 367
 Db 629 AGGATGTGAGGATAACGATCACTTTAAACAAATGAACTTCCACCTACATCACCACCTTGT 688
 QY 368 TGCAGTCTCTATAATATTATTTTTCAGGAGGCTTTTGAAAAATCATGAATTTGCACAAATG 427
 Db 689 TGCAGTCTCTATAATATTATTTTTCAGGAGGCTTTTGAAAAATCATGAATTTGCACAAATG 748
 QY 428 GACGAAATATTATAACCCAAATGACCCCAATTCATATTTCCAAGTCACAGGTTTGTGATTT 487
 Db 749 GACGAAATATTATAACCCAAATGACCCCAATTCATATTTCCAAGTCACAGGTTTGTGATTT 808
 QY 488 GGCCTGGCTTCACCTACTTCCATCCTTCAGTATGAAAAACAGCATCATCCTCTGCACTGACG 547

Db 809 GGCTGGCTTCACTACTTCCATCTTCACTATGATGAAACAGCATCATGCTCTGCACTGAGC 868
Qy 548 TTAGCCATAAAGTCCCTTCSAAGTGAGACTGTTTGGATTTCATGTTCAACTTTTATCATC 607
Db 869 TTAGCCATAAAGTCCCTTCSAAGTGAGACTGTTTGGATTTCATGTTCAACTTTTATCATC 928
Qy 608 AGACAGAAACATAAATTTCAAGACAAGATTTCACAAAGAACTAAATAGGTTAGTTGTTG 667
Db 929 AGACAGAAACATAAATTTCAAGACAAGATTTCACAAAGAACTAAATAGGTTAGTTGTTG 988
Qy 668 TTACCAAGTATAACATAAGACATACAGAGTGATGATATTGATCGGACCAAGTCCCA 727
Db 989 TTACCAAGTATAACATAAGACATACAGAGTGATGATATTGATCGGACCAAGTCCCA 1048
Qy 728 AGACACCTTTAAAGAAAGCCGAGGCTCTGAAGTCAGCTTCTTGAATATCTACAGGAAGC 787
Db 1049 AGACACCTTTAAAGAAAGCCGAGGCTCTGAAGTCAGCTTCTTGAATATCTACAGGAAGC 1108
Qy 788 AATAACAACAGAGATCACCGACTTGAAGCAGCCTGCTTTGGTCAGCCAGCCCAAGAGAA 847
Db 1109 AATAACAACAGAGATCACCGACTTGAAGCAGCCTGCTTTGGTCAGCCAGCCCAAGAGAA 1168
Qy 848 GCGGGGCGCTGGGGGACACTGCCAGGGCTGCCATGCTCATCTGAGCTGCTGCTATC 907
Db 1169 GCGGGGCGCTGGGGGACACTGCCAGGGCTGCCATGCTCATCTGAGCTGCTGCTATC 1228
Qy 908 TTACAGGCTTAACCTGATAAATCGTAATGATTTAAACGTCGATGAAAGACTTACGCGGTC 967
Db 1229 TTACAGGCTTAACCTGATAAATCGTAATGATTTAAACGTCGATGAAAGACTTACGCGGTC 1288
Qy 968 ATACAGACTTAACCTCAGAGCAAGGAGGCTGAAGTGGAGCACTCATTTGATTAATTC 1027
Db 1289 ATACAGACTTAACCTCAGAGCAAGGAGGCTGAAGTGGAGCACTCATTTGATTAATTC 1348
Qy 1028 ATAAAAACGATAATGTTCAAAGGAGCTTCGAGACTGGGGTTTGGCTTTGATTTCCAACT 1087
Db 1349 ATAAAAACGATAATGTTCAAAGGAGCTTCGAGACTGGGGTTTGGCTTTGATTTCCAACT 1408
Qy 1088 TACTGCTCTTCTCAGAGAAATTTGCAACAGAAAGATTCACCAAGGTGGAAAAACAT 1147
Db 1409 TACTGCTCTTCTCAGAGAAATTTGCAACAGAAAGATTCACCAAGGTGGAAAAACAT 1468
Qy 1148 TTGATTACATCACAAATTTGCAAGTTGTCGCAAGAAACAGAGGTGCACCATTAATTA 1207
Db 1469 TTGATTACATCACAAATTTGCAAGTTGTCGCAAGAAACAGAGGTGCACCATTAATTA 1528
Qy 1208 GTGTTAAGCCACTAGATAACTGGCTGTGATCTATACGGGAAGAAATTTAGAGCAGCCA 1267
Db 1529 GTGTTAAGCCACTAGATAACTGGCTGTGATCTATACGGGAAGAAATTTAGAGCAGCCA 1588
Qy 1268 ATTCATTGATACAAATCTATTTAAAGTTACACAGCCATGGGCATGCAAAATGAGAAAAAG 1327
Db 1589 ATTCATTGATACAAATCTATTTAAAGTTACACAGCCATGGGCATGCAAAATGAGAAAAAG 1648
Qy 1328 CAATATGATTGAAGTGGATGACAGAACTGAAGCTTACTTTAGAGTCTTACAGCAAAAG 1387
Db 1649 CAATATGATTGAAGTGGATGACAGAACTGAAGCTTACTTTAGAGTCTTACAGCAAAAG 1708
Qy 1388 TCACAGCAGACACCCAGATAGTTGCTGTGTGTGTCAGTAACTCGGAAGGCAAAATACG 1447
Db 1709 TCACAGCAGACACCCAGATAGTTGCTGTGTGTGTCAGTAACTCGGAAGGCAAAATACG 1768
Qy 1448 ATGCTATTAAAAATACCTGTGTACAGATTGCCATACCCCAAGTCAAGTGTGGTGGCCC 1507
Db 1769 ATGCTATTAAAAATACCTGTGTACAGATTGCCATACCCCAAGTCAAGTGTGGTGGCCC 1828
Qy 1508 GAACCTTAGCAACAGCAAACTGTATGCGCAATGCTACAAAGATGCCCCTACAGATGA 1567
Db 1829 GAACCTTAGCAACAGCAAACTGTATGCGCAATGCTACAAAGATGCCCCTACAGATGA 1888
Qy 1568 ACTGCAAGTGGAGGAGAGCTCTGGAGGGTGACATCCCCCTGAAGCTCGTGATCATCG 1627
Db 1889 ACTGCAAGTGGAGGAGAGCTCTGGAGGGTGACATCCCCCTGAAGCTCGTGATCATCG 1948

Qy 1628 TTGGCATCGATTGTTTACCATGACATGACAGCTGGCGGAGGTCAATCGCAGGATTTGTTG 1687
Db 1949 TTGGCATCGATTGTTTACCATGACATGACAGCTGGCGGAGGTCAATCGCAGGATTTGTTG 2008
Qy 1688 CAGCATCAATGAAGGATGACCCGCTGGTTCCTACCGTGCATATTTTCHAGGATAGAGGAC 1747
Db 2009 CAGCATCAATGAAGGATGACCCGCTGGTTCCTACCGTGCATATTTTCHAGGATAGAGGAC 2068
Qy 1748 AGGAGCTGTAGATGGGCTCAAGTCTGCTGCAAGCGCTCTGAGGGCTTGGATAGCT 1807
Db 2069 AGGAGCTGTAGATGGGCTCAAGTCTGCTGCAAGCGCTCTGAGGGCTTGGATAGCT 2128
Qy 1808 GCAATGAGTACATGCGCCGAGATCATGCTGCTACCGGATGCGCTAGGAGCGGCCAGC 1867
Db 2129 GCAATGAGTACATGCGCCGAGATCATGCTGCTACCGGATGCGCTAGGAGCGGCCAGC 2188
Qy 1868 TGAAAACTCTGTGTAACCTAGAGTGCCAGAGTGGCTTTTGGATTGTTCTAAATCCATTTGTA 1927
Db 2189 TGAAAACTCTGTGTAACCTAGAGTGCCAGAGTGGCTTTTGGATTGTTCTAAATCCATTTGTA 2248
Qy 1928 GAGGTTACAAACCTAGAGTACCGGTAATTTGGTGAAGAAAGAGTGACACACAGATTTT 1987
Db 2249 GAGGTTACAAACCTAGAGTACCGGTAATTTGGTGAAGAAAGAGTGACACACAGATTTT 2308
Qy 1988 TTGCTCAGTCTGGAGAAAGACTTCAGAATTCACATTCCTGGAAACAGTTATTGATGATAGAG 2047
Db 2309 TTGCTCAGTCTGGAGAAAGACTTCAGAATTCACATTCCTGGAAACAGTTATTGATGATAGAG 2368
Qy 2048 TTACAGACACAGATGGTATGACTTTTATTCGTGAGCAGGCTGTGAGAAGTGGTAGTG 2107
Db 2369 TTACAGACACAGATGGTATGACTTTTATTCGTGAGCAGGCTGTGAGAAGTGGTAGTG 2428
Qy 2108 TTCTCCCAACATTAATGATGATCTATGACAAAGCGGCTGAAGCCAGACCAATAC 2167
Db 2429 TTCTCCCAACATTAATGATGATCTATGACAAAGCGGCTGAAGCCAGACCAATAC 2488
Qy 2168 AGCGTGTGACCTTACAGCTGTGCCATCTATTACAACTGGCCAGGCTGTCAATCGGTTC 2227
Db 2489 AGCGTGTGACCTTACAGCTGTGCCATCTATTACAACTGGCCAGGCTGTCAATCGGTTC 2548
Qy 2228 CTGCTCTTGGCAGTACGCCCAAGCTGGCTTTCTTGTGGCCAGAGTATTACAGAG 2287
Db 2549 CTGCTCTTGGCAGTACGCCCAAGCTGGCTTTCTTGTGGCCAGAGTATTACAGAG 2608
Qy 2288 AGCCAAATCTGTCACTGTCAACCGCTTTTACTACTCTAA 2328
Db 2609 AGCCAAATCTGTCACTGTCAACCGCTTTTACTACTCTAA 2649

RESULT 3
AF104260
LOCUS
DEFINITION
ACCESSION
VERSION
KEYWORDS
SOURCE
ORGANISM
REFERENCE
AUTHORS
TITLE
JOURNAL
MEDLINE
PUBMED
REFERENCE
AUTHORS
TITLE
JOURNAL

AF104260 3362 bp mRNA linear PRI 30-APR-2003
Homo sapiens HIWI mRNA, complete cds.
AF104260
AF104260.2 GI:18098557
Homo sapiens (human)
Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 3362)
Cox,D.N., Chao,A., Baker,J., Chang,L., Qiao,D. and Lin,H.
A novel class of evolutionarily conserved genes defined by piwi are
essential for stem cell self-renewal
Genes Dev. 12 (23), 3715-3727 (1998)
99069219
9851978
2 (bases 1 to 3362)
Qiao,D., Zeeman,A.-M., Deng,W., Looijenga,L.H.J. and Lin,H.
Molecular characterization of hiwi, a human member of the piwi gene
family whose overexpression is correlated to seminomas
Oncogene 21 (25), 3988-3999 (2002)

```

MEDLINE 22033724
PUBMED 12037681
REFERENCE 3 (bases 1 to 3362)
AUTHORS Lin,H.
TITLE Direct Submission
JOURNAL Submitted (04-NOV-1998) Cell Biology, Duke University Medical
Center, 412 Nanaline Duke Bldg., Research Dr., Durham, NC 27710,
USA
REFERENCE 4 (bases 1 to 3362)
AUTHORS Qiao,D., Zeeman,A.-M., Deng,W., Looijenga,L.H.J. and Lin,H.
TITLE Direct Submission
JOURNAL Submitted (09-JAN-2002) Cell Biology, Duke University, 412 Nanaline
Duke Building, DUMC, Durham, NC 27710, USA
REMARK Nucleotide sequence updated by submitter
COMMENT On Jan 9, 2002 this sequence version replaced gi:4038412.
FEATURES
    Location/Qualifiers
        1..3362
            /organism="Homo sapiens"
            /mol_type="mRNA"
            /db_xref="taxon:9606"
            /chromosome="12"
            /map="12q24.33"
            /tissue_type="testis"
            /note="member of the PIWI family"
            /codon_start=1
            /product="HIWI"
            /protein_id="AAC97371.2"
            /db_xref="GI:18098558"
            /translation="MTGRARARARGRQGETAQLVGSTASQPGYIOPRPQPPAEG
            ELFGRRQRTAGGTAKSGLOI SAGFQSLAERGRRRDFDLGVNTRQNLHVKE
            SXTGSSGIVLSTVNHFRITSPOWALOVYHIDYNEPLMEARLRGALLFOHEDLIGKC
            HAPDTGILPKRLQKQVTEVSKTNGEDVRITLTNELPPTSCLOFYNIIFRR
            LUKIMLOQIGRYNINPNDPIDPSRLVINPFGFTISILQYENSIMLCTDVSHKVLRS
            ETVLDFMFYHTEHSEKIQEQLIGLVLTIKYNIKRVDDIDWDONPKSTFKK
            ADGSGVSEFYRKYQNYEITDLKQPLVSPKRRPGGTLPGPAMLIPELCYLTGL
            TDKMRNDFNMKDLAVHTLTPQOREVGRILDIYHKNDVQRELRDWGLSFDNLL
            SPSGR LQFETKHOGKTPDYNPQADWSKETRGAPLISVKPLDNWLLLYTRNVEAA
            NSLIQNLKFTPMKMGOMSKAJMIEVDVDETEAVLRVLOQVTDATQIVCLLSNPKO
            KYDAIKYICTDPTSCQVARTLKGQTVMAIAIKALQNKCMGSLMRVDIPLK
            LVNIVGIDCYDHMTAGRSIAGVINSNEGTRWFRFCIFQDRQBELVDGLKVCLOAA
            LKRWNSCYMPSRIITVYDGVGQGLKTLVNYEVPQFLDKLSIRGRYNPLRTIVV
            KRVNTRFPAQSGRLNDELPGTVIDVETREPWDFITVSOAVRSGSVSPHYNVIY
            DNSLKPDPHIOQLRYKLCHYYNWPVGIVRPAPQVAVHKLAFIVGQS IHREPNLSLN
            RLYYL"
BASE COUNT 1026 a 659 c 759 g 918 t
ORIGIN
Query Match 99.4%; Score 2313; DB 9; Length 3362;
Best Local Similarity 99.8%; Pred. No. 0;
Matches 2316; Conservative 0; Mismatches 5; Indels 0; Gaps 0;
QY 8 TTGGTGTGAACAAGCGAGAACTTAGACCATGTTAAAGAAATCAAAAACAGGTTCTTCAG 67
DB 322 TTGGTGTGAATACAAGCGAGAACTTAGACCATGTTAAAGAAATCAAAAACAGGTTCTTCAG 381
QY 68 GCATTATAGTAAGGTTAAGCACTAACCATTTCCGGCTGCATCCCGTCCCGAGTGGSCCT 127
DB 382 GCATTATAGTAAGGTTAAGCACTAACCATTTCCGGCTGCATCCCGTCCCGAGTGGSCCT 441
QY 128 TATATCAGTATCATGATGACTATATACCCACTGATGGAAGCCAGAGAACTCCGTTCCAGTCT 187
DB 442 TATATCAGTATCATGATGACTATATACCCACTGATGGAAGCCAGAGAACTCCGTTCCAGTCT 501
QY 188 TTCTTTTCAACGAGAGATCTAATTGGAAAGTGTGATGCTTTTGTAGTGAACCATATAT 247
DB 502 TTCTTTTCAACGAGAGATCTAATTGGAAAGTGTGATGCTTTTGTAGTGAACCATATAT 561
QY 248 TTTTACCTTAAAGACTACAGCAAAAGGTTACTGAGTTTGTAGTGAACCGGAAATGGAG 307
DB 562 TTTTACCTTAAAGACTACAGCAAAAGGTTACTGAGTTTGTAGTGAACCGGAAATGGAG 621
QY 308 AGGATGTGAGGATAACGATCACTTTAACAAATGAAGCTTCCACCTACATCAACCACTTGT 367

```

```

622 AGGATGTGAGGATAACGATCACTTTAAACAATGAAGCTTCCACCTACATCAACCACTTGT 681
QY 368 TGCAGTCTTAATAATATATATTTTCAGGAGGCTTTGAAAATCATGAATTTGCAACAATTTG 427
DB 682 TGCAGTCTTAATAATATATATTTTCAGGAGGCTTTGAAAATCATGAATTTGCAACAATTTG 741
QY 428 GACGAAATTTATTAACCAATGACCAATGATATTCCAAAGTCAACAGGTTGGTGAATTT 487
DB 742 GACGAAATTTATTAACCAATGACCAATGATATTCCAAAGTCAACAGGTTGGTGAATTT 801
QY 488 GGCCTGGCTTCACCTACTTCCATTCCTAGTATGAAAAAGCATCATGCTCTGCATGACG 547
DB 802 GGCCTGGCTTCACCTACTTCCATTCCTAGTATGAAAAAGCATCATGCTCTGCATGACG 861
QY 548 TTAGCCATAAAGTCTTCGAAAGTGAAGTGTGTTGGAATTCATGTTCACTTTTATCATC 607
DB 862 TTAGCCATAAAGTCTTCGAAAGTGAAGTGTGTTGGAATTCATGTTCACTTTTATCATC 921
QY 608 AGACAGAAAGAACATAAATTTCAAGAAACAAAGTTTCCAAAGAACTAATAGTTTGTGTTTC 667
DB 922 AGACAGAAAGAACATAAATTTCAAGAAACAAAGTTTCCAAAGAACTAATAGTTTGTGTTTC 981
QY 668 TTACCAAGTATAACAATAAGACATACAGAGTGGATGATATTGACTGGGACCGAATATCCCA 727
DB 982 TTACCAAGTATAACAATAAGACATACAGAGTGGATGATATTGACTGGGACCGAATATCCCA 1041
QY 728 AGACACACCTTTAAGAAAGCCGACGCTCTGAAGTCAAGTCTTCTTAGAATACACAGGAGC 787
DB 1042 AGACACACCTTTAAGAAAGCCGACGCTCTGAAGTCAAGTCTTCTTAGAATACACAGGAGC 1101
QY 788 AATACAAACCAAGAGATCAACGACTTTGAAGACAGCCTGTCTTGCTGACGACGACCAAGAGAA 847
DB 1102 AATACAAACCAAGAGATCAACGACTTTGAAGACAGCCTGTCTTGCTGACGACGACCAAGAGAA 1161
QY 848 GCGCGGGCCCTGGGGGGGACATGCGCAGGCGCTGCCATGCTCTTCTGAGCTCTGCTATC 907
DB 1162 GCGCGGGCCCTGGGGGGGACATGCGCAGGCGCTGCCATGCTCTTCTGAGCTCTGCTATC 1221
QY 908 TTACAGGCTTAACCTGATTAATGCGTAAATGATTTTAAAGTGAAGAACTTAGCGGTTTC 967
DB 1222 TTACAGGCTTAACCTGATTAATGCGTAAATGATTTTAAAGTGAAGAACTTAGCGGTTTC 1281
QY 968 ATACAGACTTAACCTCCAGAGCAAGAGCAGCGTGAAGTGGGACGACTCATTTGATTAATTC 1027
DB 1282 ATACAGACTTAACCTCCAGAGCAAGAGCAGCGTGAAGTGGGACGACTCATTTGATTAATTC 1341
QY 1028 ATAAAAACGATAAATGTTCAAGGAGCTTCGAGACTGGGGTTTGAGCTTTGATTTCAACT 1087
DB 1342 ATAAAAACGATAAATGTTCAAGGAGCTTCGAGACTGGGGTTTGAGCTTTGATTTCAACT 1401
QY 1088 TACTGTCCTTCTCAGGAGAAATTTTGCACCAAGAAAGATTTCCCAAGTGGGAAACAT 1147
DB 1402 TACTGTCCTTCTCAGGAGAAATTTTGCACCAAGAAAGATTTCCCAAGTGGGAAACAT 1461
QY 1148 TTGATTACAATCCAAATTTGAGATTTGGTCCAAAGAAACAAAGAGTGCACCATTAATTA 1207
DB 1462 TTGATTACAATCCAAATTTGAGATTTGGTCCAAAGAAACAAAGAGTGCACCATTAATTA 1521
QY 1208 GTGTTAAGCCACTAGATAACTGGCTGTGATCTATACGCGAAAGAAATTTAGAGAGCCCA 1267
DB 1522 GTGTTAAGCCACTAGATAACTGGCTGTGATCTATACGCGAAAGAAATTTAGAGAGCCCA 1581
QY 1268 ATTCATTGATACAAATCTTATTTAAGTTACACCGCCATGGCATGCAAAATGAGAAAG 1327
DB 1582 ATTCATTGATACAAATCTTATTTAAGTTACACCGCCATGGCATGCAAAATGAGAAAG 1641
QY 1328 CAATATGATTGAAGTGAATGACAGAACTGAAGCCCTACTTAAAGAGTCTTACAGCAAAAG 1387
DB 1642 CAATATGATTGAAGTGAATGACAGAACTGAAGCCCTACTTAAAGAGTCTTACAGCAAAAG 1701
QY 1388 TCACAGCAGACACCCAGATAGTTGTCTGTCTGTCTCAAGTAAATCGGAAGGACCAATACG 1447

```


QY 2288 AGCCAAATCTGCTCAACCGGCTTTACTACCTCTAA 2328
Db 2619 AGCCAAATCTGCTCAACCGGCTTTACTACCTCTAA 2659

RESULT 5
AK093133
LOCUS
DEFINITION Homo sapiens cDNA FLJ35814 fis, clone TEST12006083, highly similar to Mus musculus mRNA for MIWI (piwi).
ACCESSION AK093133
VERSION 1 GI:21751907
KEYWORDS oligo capping; fis (full insert sequence).
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens

REFERENCE
AUTHORS Oshima, A., Takahashi-Fujii, A., Tanase, T., Imose, N., Takeuchi, K., Arita, M., Musashino, K., Yuuki, H., Hara, H., Sugiyama, T., Irie, R., Otsuki, T., Sato, H., Wakamatsu, A., Ishii, S., Yamamoto, J., Isono, Y., Kawai-Hio, Y., Saito, K., Nishikawa, T., Kimura, K., Yamashita, H., Matsuo, K., Nakamura, Y., Sekine, M., Kikuchi, H., Kanda, K., Wagatsuma, M., Murakawa, K., Kanehori, K., Sugiyama, A., Kawakami, B., Suzuki, Y., Sugano, S., Nagahara, K., Masuho, Y., Nagai, K. and Isogai, T.
1
NEDO human cDNA sequencing project
Unpublished
2 (bases 1 to 2579)
Isogai, T. and Yamamoto, J.
Direct Submission
Submitted (04-JUL-2002) Takao Isogai, FLJ Project (HRI Team); 2-6-7
Kazusa-Kamatari, Kisarazu, Chiba 292-0812, Japan
(E-mail: genomics@hri.co.jp, Tel: 81-438-52-3975, Fax: 81-438-52-3986)
NEDO human cDNA sequencing project supported by Ministry of Economy, Trade and Industry of Japan; cDNA full insert sequencing: Research Association for Biotechnology (RAB); cDNA library construction: Helix Research Institute (HRI) (supported by Japan Key Technology Center etc.); 5'- & 3'-end one pass sequencing: RAB, HRI, and Biotechnology Center, National Institute of Technology and Evaluation; clone selection for full insert sequencing: HRI and RAB; annotation: HRI and RAB.

LOCATION/Qualifiers
1. .2579
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="TEST12006083"
/tissue_type="testis"
/clone_lib="TEST12"
/notes="Cloning vector: pME18SFL3"
54. .2543
/notes="unnamed protein product"
/codon_start=1
/protein_id="BAC04068.1"
/db_xref="GI:21751908"
/translation="MTGRARARARGARQETALVNSTASQDPYIQRPPQPPRAG
ELFGRQRTAGTAKSLQISAGFQELSLAERGRRRRDFDLGNTRQNLHDVKE
SKTSSGIIIVRLSTNFRITSFPQWALYQYHIDYNELMEARLRSLALLGHEDLIKC
HAPDQILFLPRLQKQKATEVSKTRNGEDVRITITLNEPPTSPTCLQFYNIIPRR
LKIMLQOIGRNYNPNDPIDPSHRLVTPGFTTSILQYENSIMLCTDVSHKVLRS
ETVLDFENFYHCTEHKQEQVSKELIGLVILTKNNKYRYRVDLDWQNPSPFKK
ADGSEVSFLYRKQNYKEITDQPVLSVQPKRRPGGTLPGPAMLIPELCYLTGL
TDKMRDNFNMQLAVHTLITDQPVLSVQPKRRPGGTLPGPAMLIPELCYLTGL
SFSGRITQETKIHQKGTDPYNQADWSKETRGAPLISVKPLDNLMLLYTRNRYEA
NSLIQNFQKTPANGQMRKAIMIEVDRTAYLRVLOQKVTADTOIVVCLLSNRKD
KYDAIKYLCCTPSCQVAFELCKQQTWMAIKIALQMNCKWGGELNWDIPLK
LWMLVIGDCHEMTAGRSSTAGFVASINEGMTWFRSCIFQDRGQSLDGLKVCLOQA
LRWNSCNETMPSRIIVIRVDGVDGQKLTIVNVEVPFLDCLKISIGRGNPRITV
KRGVNTFRFAQSGRLQNPETGVIVDEVTREPEWYDFIVSQVRSGSVSPHYNYI
DNSGLKPDHIQRIYKLCIHYNNWVSASTC"

REFERENCE
AUTHORS Oshima, A., Takahashi-Fujii, A., Tanase, T., Imose, N., Takeuchi, K., Arita, M., Musashino, K., Yuuki, H., Hara, H., Sugiyama, T., Irie, R., Otsuki, T., Sato, H., Wakamatsu, A., Ishii, S., Yamamoto, J., Isono, Y., Kawai-Hio, Y., Saito, K., Nishikawa, T., Kimura, K., Yamashita, H., Matsuo, K., Nakamura, Y., Sekine, M., Kikuchi, H., Kanda, K., Wagatsuma, M., Murakawa, K., Kanehori, K., Sugiyama, A., Kawakami, B., Suzuki, Y., Sugano, S., Nagahara, K., Masuho, Y., Nagai, K. and Isogai, T.
1
NEDO human cDNA sequencing project
Unpublished
2 (bases 1 to 2579)
Isogai, T. and Yamamoto, J.
Direct Submission
Submitted (04-JUL-2002) Takao Isogai, FLJ Project (HRI Team); 2-6-7
Kazusa-Kamatari, Kisarazu, Chiba 292-0812, Japan
(E-mail: genomics@hri.co.jp, Tel: 81-438-52-3975, Fax: 81-438-52-3986)
NEDO human cDNA sequencing project supported by Ministry of Economy, Trade and Industry of Japan; cDNA full insert sequencing: Research Association for Biotechnology (RAB); cDNA library construction: Helix Research Institute (HRI) (supported by Japan Key Technology Center etc.); 5'- & 3'-end one pass sequencing: RAB, HRI, and Biotechnology Center, National Institute of Technology and Evaluation; clone selection for full insert sequencing: HRI and RAB; annotation: HRI and RAB.

LOCATION/Qualifiers
1. .2579
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="TEST12006083"
/tissue_type="testis"
/clone_lib="TEST12"
/notes="Cloning vector: pME18SFL3"
54. .2543
/notes="unnamed protein product"
/codon_start=1
/protein_id="BAC04068.1"
/db_xref="GI:21751908"
/translation="MTGRARARARGARQETALVNSTASQDPYIQRPPQPPRAG
ELFGRQRTAGTAKSLQISAGFQELSLAERGRRRRDFDLGNTRQNLHDVKE
SKTSSGIIIVRLSTNFRITSFPQWALYQYHIDYNELMEARLRSLALLGHEDLIKC
HAPDQILFLPRLQKQKATEVSKTRNGEDVRITITLNEPPTSPTCLQFYNIIPRR
LKIMLQOIGRNYNPNDPIDPSHRLVTPGFTTSILQYENSIMLCTDVSHKVLRS
ETVLDFENFYHCTEHKQEQVSKELIGLVILTKNNKYRYRVDLDWQNPSPFKK
ADGSEVSFLYRKQNYKEITDQPVLSVQPKRRPGGTLPGPAMLIPELCYLTGL
TDKMRDNFNMQLAVHTLITDQPVLSVQPKRRPGGTLPGPAMLIPELCYLTGL
SFSGRITQETKIHQKGTDPYNQADWSKETRGAPLISVKPLDNLMLLYTRNRYEA
NSLIQNFQKTPANGQMRKAIMIEVDRTAYLRVLOQKVTADTOIVVCLLSNRKD
KYDAIKYLCCTPSCQVAFELCKQQTWMAIKIALQMNCKWGGELNWDIPLK
LWMLVIGDCHEMTAGRSSTAGFVASINEGMTWFRSCIFQDRGQSLDGLKVCLOQA
LRWNSCNETMPSRIIVIRVDGVDGQKLTIVNVEVPFLDCLKISIGRGNPRITV
KRGVNTFRFAQSGRLQNPETGVIVDEVTREPEWYDFIVSQVRSGSVSPHYNYI
DNSGLKPDHIQRIYKLCIHYNNWVSASTC"

FEATURES
source
1. .2579
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="TEST12006083"
/tissue_type="testis"
/clone_lib="TEST12"
/notes="Cloning vector: pME18SFL3"
54. .2543
/notes="unnamed protein product"
/codon_start=1
/protein_id="BAC04068.1"
/db_xref="GI:21751908"
/translation="MTGRARARARGARQETALVNSTASQDPYIQRPPQPPRAG
ELFGRQRTAGTAKSLQISAGFQELSLAERGRRRRDFDLGNTRQNLHDVKE
SKTSSGIIIVRLSTNFRITSFPQWALYQYHIDYNELMEARLRSLALLGHEDLIKC
HAPDQILFLPRLQKQKATEVSKTRNGEDVRITITLNEPPTSPTCLQFYNIIPRR
LKIMLQOIGRNYNPNDPIDPSHRLVTPGFTTSILQYENSIMLCTDVSHKVLRS
ETVLDFENFYHCTEHKQEQVSKELIGLVILTKNNKYRYRVDLDWQNPSPFKK
ADGSEVSFLYRKQNYKEITDQPVLSVQPKRRPGGTLPGPAMLIPELCYLTGL
TDKMRDNFNMQLAVHTLITDQPVLSVQPKRRPGGTLPGPAMLIPELCYLTGL
SFSGRITQETKIHQKGTDPYNQADWSKETRGAPLISVKPLDNLMLLYTRNRYEA
NSLIQNFQKTPANGQMRKAIMIEVDRTAYLRVLOQKVTADTOIVVCLLSNRKD
KYDAIKYLCCTPSCQVAFELCKQQTWMAIKIALQMNCKWGGELNWDIPLK
LWMLVIGDCHEMTAGRSSTAGFVASINEGMTWFRSCIFQDRGQSLDGLKVCLOQA
LRWNSCNETMPSRIIVIRVDGVDGQKLTIVNVEVPFLDCLKISIGRGNPRITV
KRGVNTFRFAQSGRLQNPETGVIVDEVTREPEWYDFIVSQVRSGSVSPHYNYI
DNSGLKPDHIQRIYKLCIHYNNWVSASTC"

CDS
1. .2543
/notes="unnamed protein product"
/codon_start=1
/protein_id="BAC04068.1"
/db_xref="GI:21751908"
/translation="MTGRARARARGARQETALVNSTASQDPYIQRPPQPPRAG
ELFGRQRTAGTAKSLQISAGFQELSLAERGRRRRDFDLGNTRQNLHDVKE
SKTSSGIIIVRLSTNFRITSFPQWALYQYHIDYNELMEARLRSLALLGHEDLIKC
HAPDQILFLPRLQKQKATEVSKTRNGEDVRITITLNEPPTSPTCLQFYNIIPRR
LKIMLQOIGRNYNPNDPIDPSHRLVTPGFTTSILQYENSIMLCTDVSHKVLRS
ETVLDFENFYHCTEHKQEQVSKELIGLVILTKNNKYRYRVDLDWQNPSPFKK
ADGSEVSFLYRKQNYKEITDQPVLSVQPKRRPGGTLPGPAMLIPELCYLTGL
TDKMRDNFNMQLAVHTLITDQPVLSVQPKRRPGGTLPGPAMLIPELCYLTGL
SFSGRITQETKIHQKGTDPYNQADWSKETRGAPLISVKPLDNLMLLYTRNRYEA
NSLIQNFQKTPANGQMRKAIMIEVDRTAYLRVLOQKVTADTOIVVCLLSNRKD
KYDAIKYLCCTPSCQVAFELCKQQTWMAIKIALQMNCKWGGELNWDIPLK
LWMLVIGDCHEMTAGRSSTAGFVASINEGMTWFRSCIFQDRGQSLDGLKVCLOQA
LRWNSCNETMPSRIIVIRVDGVDGQKLTIVNVEVPFLDCLKISIGRGNPRITV
KRGVNTFRFAQSGRLQNPETGVIVDEVTREPEWYDFIVSQVRSGSVSPHYNYI
DNSGLKPDHIQRIYKLCIHYNNWVSASTC"

BASE COUNT
ORIGIN
797 a 542 c 613 g 627 t

Query Match 94.5%; Score 2200.2; DB 9; Length 2579;
Best Local Similarity 99.9%; Pred. No. 0;
Matches 2202; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 8 TTGGTGTGACACAGGAGAACTTTAGACCATGTTAAAGAAATCAAAGAGGTTCTTCAG 67
Db 319 TTGGTGTGATTAACAGGAGAACTTTAGACCATGTTAAAGAAATCAAAGAGGTTCTTCAG 378
QY 68 GCATTATAGTAGGTTAAGCACTAACCACTTCCGGCTGACATCCCGTCCCGAGTGGCCCT 127
Db 379 GCATTATAGTAGGTTAAGCACTAACCACTTCCGGCTGACATCCCGTCCCGAGTGGCCCT 438
QY 128 TATATCAGTATCACAATTGACTATATACCACTGATGGAAGCCAGAGACTCCGTTGAGTTC 187
Db 439 TATATCAGTATCACAATTGACTATATACCACTGATGGAAGCCAGAGACTCCGTTGAGTTC 498
QY 188 TTCTTTTTCACACAGGAGAACTTTAAGTGAAGTGTGATGCTTTTTCATGGAAGCATATTAT 247
Db 499 TTCTTTTTCACACAGGAGAACTTTAAGTGAAGTGTGATGCTTTTTCATGGAAGCATATTAT 558
QY 248 TTTTACCTTAAAGAGACTACAGCAAAAAGGTTACTGAAAGTCTTTAGTAAAGACCCGGAATGGAG 307
Db 559 TTTTACCTTAAAGAGACTACAGCAAAAAGGTTACTGAAAGTCTTTAGTAAAGACCCGGAATGGAG 618
QY 308 AGGATGTGAGGATAACGATCACTTTTAAACAATGAACTTCCACCTACATCAACAATTTGTT 367
Db 619 AGGATGTGAGGATAACGATCACTTTTAAACAATGAACTTCCACCTACATCAACAATTTGTT 678
QY 368 TGCAGTCTTATAATATATTTTCAGAGGCTTTTCAAGAAATCATGATTTGCAACAATTTG 427
Db 679 TGCAGTCTTATAATATATTTTCAGAGGCTTTTCAAGAAATCATGATTTGCAACAATTTG 738
QY 428 GACGAAATTTATATACCAATGACCAATGATTTTCCAAAGTACACAGTTGGTGATTT 487
Db 739 GACGAAATTTATATACCAATGACCAATGATTTTCCAAAGTACACAGTTGGTGATTT 798
QY 488 GCGCTGGCTTCACTACTTCCATCTTCACTAGTATGAAAGACAGCATGCTCTGCACTGACG 547
Db 799 GCGCTGGCTTCACTACTTCCATCTTCACTAGTATGAAAGACAGCATGCTCTGCACTGACG 858
QY 548 TTAGCCATAAAGTCTTCCGAGTGAGACTGTTTTCGATTTTCATGTTTCACTTTTATCATC 607
Db 859 TTAGCCATAAAGTCTTCCGAGTGAGACTGTTTTCGATTTTCATGTTTCACTTTTATCATC 918
QY 608 AGACAGAGACACATAAAATTTCAAGAACCAAGTTTCCAAAGAACTAATAGTGTGTTTC 667
Db 919 AGACAGAGAGACATAAAATTTCAAGAACCAAGTTTCCAAAGAACTAATAGTGTGTTTC 978
QY 668 TTACCAAGTATTAACAATAAGACATACAGATGGATGATTTGACTGGGACCGAATCCCA 727
Db 979 TTACCAAGTATTAACAATAAGACATACAGATGGATGATTTGACTGGGACCGAATCCCA 1038
QY 728 AGAGCACCTTTAAGAAAGCCGACGCTCTGAAGTCAAGCTTTTGAAGTACTACAGGAAGC 787
Db 1039 AGAGCACCTTTAAGAAAGCCGACGCTCTGAAGTCAAGCTTTTGAAGTACTACAGGAAGC 1098
QY 788 AATAACAACCAAGAGATCAACCGCTTTGAAGACGCTGTCTTGTGTCAGCCACCCCAAGAGAA 847
Db 1099 AATAACAACCAAGAGATCAACCGCTTTGAAGACGCTGTCTTGTGTCAGCCACCCCAAGAGAA 1158
QY 848 GCGGGGGCCCTGGGGGGACATGCGCAGGCGCTGCCATGCTATTCTCTGAGCTCTGCTATC 907
Db 1159 GCGGGGGCCCTGGGGGGACATGCGCAGGCGCTGCCATGCTATTCTCTGAGCTCTGCTATC 1218
QY 908 TTACAGGCTCTAATCATAAATGCGTAATGATTTTAAAGTGTGATGAAGACTTAGCGGTTTC 967
Db 1219 TTACAGGCTCTAATCATAAATGCGTAATGATTTTAAAGTGTGATGAAGACTTAGCGGTTTC 1278
QY 968 ATACAAGACTTAATCCAGAGCAAAAGCGCTGAAGTGGGACGACTCATTGATTACATTTC 1027
Db 1279 ATACAAGACTTAATCCAGAGCAAAAGCGCTGAAGTGGGACGACTCATTGATTACATTTC 1338

```
1028 ATAAAAACGATATGTTCAAAGGAGCTTCGAGACTGGGGTTTGAGCTTTGATTCACACT 1087
1339 ATAAAAACGATATGTTCAAAGGAGCTTCGAGACTGGGGTTTGAGCTTTGATTCACACT 1398
1088 TACTGTCTCTCAGGAAGATTTTCCAAACAGAAAGATTTCCAAAGTGGAAAAACAT 1147
1399 TACTGTCTCTCAGGAAGATTTTCCAAACAGAAAGATTTCCAAAGTGGAAAAACAT 1458
1148 TTGATTACATCCACAAATTTGCAGATTTGGTCCAAAGAAACAGAGGTGCACCATTAATTA 1207
1459 TTGATTACATCCACAAATTTGCAGATTTGGTCCAAAGAAACAGAGGTGCACCATTAATTA 1518
1208 GTGTTAAGCCACTAGATACTGGCTGTGATCTATACGCGAAGAAATTTATGAAGCAGCCA 1267
1519 GTGTTAAGCCACTAGATACTGGCTGTGATCTATACGCGAAGAAATTTATGAAGCAGCCA 1578
1268 ATTCAATTGATACAAATCTATTATTAAGTTACACCGCCATGGGCATGCAAAATGAGAAAG 1327
1579 ATTCAATTGATACAAATCTATTATTAAGTTACACCGCCATGGGCATGCAAAATGAGAAAG 1638
1328 CCAATATGATTGAGTGGATGACAGAACTGAAGCCCTACTTTAGAGTCTTACAGCAAAAGG 1387
1639 CCAATATGATTGAGTGGATGACAGAACTGAAGCCCTACTTTAGAGTCTTACAGCAAAAGG 1698
1388 TCACAGCAGACACCCAGATAGTTGTCTGTCTGTCTGTCTGTCTGTCTGTCTGTCTGTCTGT 1447
1699 TCACAGCAGACACCCAGATAGTTGTCTGTCTGTCTGTCTGTCTGTCTGTCTGTCTGTCTGT 1758
1448 ATGCTATTAATAATACCTGTGTACAGATTTGCCCTACCCCAAGTCAAGTGTGGTGGCCC 1507
1759 ATGCTATTAATAATACCTGTGTACAGATTTGCCCTACCCCAAGTCAAGTGTGGTGGCCC 1818
1508 GACCTTAGCAACACGCAAACTGTATGCGCATCTGCTACAAAGATTTGCCCTACAGATGA 1567
1819 GACCTTAGCAACACGCAAACTGTATGCGCATCTGCTACAAAGATTTGCCCTACAGATGA 1878
1568 ACTGCAAGTGGAGGAGAGCTCTCGAGGTGGACATCCCTCGAAGCTCGTATGATCG 1627
1879 ACTGCAAGTGGAGGAGAGCTCTCGAGGTGGACATCCCTCGAAGCTCGTATGATCG 1938
1628 TTGGCATCGATTTGTTACCATGACATGACAGCTCGGCGGAGGTCAATCGCAGGATTTGTTG 1687
1939 TTGGCATCGATTTGTTACCATGACATGACAGCTCGGCGGAGGTCAATCGCAGGATTTGTTG 1998
1688 CAGCATCAATGAGGATGACCGCGTGTCTCAGCTGCATATTTTACAGTATGAGGAC 1747
1999 CAGCATCAATGAGGATGACCGCGTGTCTCAGCTGCATATTTTACAGTATGAGGAC 2058
1748 AGGAGCTGGTAGTGGCTCAAACTCTGCTGCAAGCGGCTCTGAGGGCTTTGGAATAGCT 1807
2059 AGGAGCTGGTAGTGGCTCAAACTCTGCTGCAAGCGGCTCTGAGGGCTTTGGAATAGCT 2118
1808 GCAATGAGTACATGCCAGCGGATCATCTGTACCGCGATGGCGTAGGAGACGGCAGC 1867
2119 GCAATGAGTACATGCCAGCGGATCATCTGTACCGCGATGGCGTAGGAGACGGCAGC 2178
1868 TGAACACATGTTGAACTAGCACTGAGTGCACAGTTTGTGTTGTTCTCTAAATCCATTTGTA 1927
2179 TGAACACATGTTGAACTAGCACTGAGTGCACAGTTTGTGTTGTTCTCTAAATCCATTTGTA 2238
1928 GAGTTACAAACCTTAGACTAACGTAATTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTT 1987
2239 GAGTTACAAACCTTAGACTAACGTAATTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTT 2298
1988 TTGCTCAGTCTGAGGAGAGCTTACAGATCCACTTCTCTGAAACAGTTTATGATGTAGAGG 2047
2299 TTGCTCAGTCTGAGGAGAGCTTACAGATCCACTTCTCTGAAACAGTTTATGATGTAGAGG 2358
2048 TTACCAACAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 2107
2359 TTACCAACAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 2418
2108 TTCTCCACACATTAATGTCATCTATGACAAACAGCGGCTTGAAGCCAGACCATAC 2167
```

```
Db 2419 TTTCTCCACACATTTACATGTCATCTATGACACAGCGCTGAAGCCAGACCATAC 2478
QY 2168 AGCGTTGACCTACAGCTGTGCCACATCTATTACAACTGGCCAG 2212
Db 2479 AGCGTTGACCTACAGCTGTGCCACATCTATTACAACTGGCCAG 2523

RESULT 6
AB032604 3910 bp mRNA linear ROD 04-OCT-2001
Mus musculus mRNA for MIWI, complete cds.
DEFINITION
ACCESSION AB032604
VERSION AB032604.1 GI:7416110
KEYWORDS MIWI (piwi). (house mouse)
SOURCE Mus musculus
ORGANISM Mus musculus
REFERENCE
1 Kuramochi-Miyagawa, S., Kimura, T., Yomogida, K., Kuroiwa, A., Tadokoro, Y., Fujita, Y., Sato, M., Matsuda, Y., and Nakano, T. Two mouse piwi-related genes: miwi and miwi Mech. Dev. 108 (1-2), 121-133 (2001)
2 Miyagawa, S.K. and Nakano, T. Direct Submission
3-1. Suita, Osaka 565-0871, Japan (E-mail: smiyagawa@biken.osaka-u.ac.jp, Tel: 81-6-6879-8362, Fax: 81-6-6879-8362)
LOCATION/Qualifiers
1. 3910
/organism="Mus musculus"
/mol_type="mRNA"
/db_xref="taxon:10090"
/map="5F distal-G2 proximal"
/tissue_lib="testis"
gene 1. 3910
CDS 156..274
/gene="miwi"
/contig="miwi"
/codon_start=1
/product="MIWI"
/protein_id="BAA93705.1"
/db_xref="GI:7416111"
/translation="MTGRARAPARGARGOETVQHVGAASOQPGYIPRPOSPTEG
DLVGRGRGMVTVGATSKSQELIAGFGQELSLAERGRRRDFHDLGVNTRQNLHVYK
BKTSGSGGLIVKLSTNHFRLTSRQWALYQYHIDYNPLMEARRLSALLFQHEDLIGR
CHAFPTGLFLFKRLQHKVTSFVSRGHRVITITLNEPLPTSCLOQYNIIFR
RLIKMNLQIIGRYNPSPDIPIDENHRLIWPGETTSILOYNIMLCTDQNPXKFLR
STVLDVFNLYQOETEEHFKQSKELIGLVLTKNNKTVRVDIDWDQNPXKFLR
KADGSEVFLVYRQVQNELIDLPQVLSOPKSRGPGGTLPGPAMLIPLCYLTG
LTDKRNDFNVKMDLAHTRLTPEQRQREVRGLIYIHKDDNVQRLDWGLSFDNL
LSFQSKILQSEKIHQGGCTFDYNFQADMSKETRGAPLISVKPLNWLIIYTRNREA
ANSLIQNFKVTTPAMQIMKKAIMEVDRTQATYALFALQKQVTSQTVIVCLLSNRK
DKYDAIKKYLCTDCTPSCVAVTLGQQTWMAITKIALQWNCMGGLWRVDMPL
KLAMTVGIDCVHDTTAGRSIAGFVASINEGRWFSRCVDFQDGLVGLKVCLOA
ALRAMGCGNEMPSVIVYRGVGGQKTLANVEVPOPLDCLKSVGRGVNPLRTVIV
VKRVNARFPAQSGRLQNLPGTVIDVEVTPEDVDFPIVQVARSQSVSPHYNVI
YDSSGLKPDHPIQLRTYKLCVYNNPWPVIRPAPCOYAKHLAPLVGQSHRPFNSLS
NELYI"
BASE COUNT 993 a 970 c 1063 g 884 t
ORIGIN
Query Match 72.4%; Score 1684.6; DB 10; Length 3910;
Best Local Similarity 82.8%; Pred. No. 0;
Matches 1924; Conservative 0; Mismatches 399; Indels 0; Gaps 0;
```



```
Db 2582 CCAGCGGCTGACATCAAGCTCTGCCACGTGTACTATATTTGGCGCTGAGTCAATCGAGT 2641
Qy 2226 TCCTCTCTCTTCCAGTACGCCCAAGCTGGCTTTCTTGTGTGCCAGAGTATTCACAG 2285
Db 2642 CCCTGCACCTTCCAGTATGACACAAGCTGGCTTCTCGTGGCCAGAGATCCACAG 2701
Qy 2286 AGAGCCAAATCTGTCACTCTCAAAACCGCTTTACTACCTCTAA 2328
Db 2702 AGAGCCAAACCTCTCCCTGTCCACCGCTCTACTACCTCTAA 2744

RESULT 7
AF438405 4064 bp mRNA linear ROD 13-NOV-2001
LOCUS Mus musculus MIWI mRNA, complete cds.
DEFINITION AF438405
ACCESSION AF438405
VERSION AF438405.1 GI:16905060
KEYWORDS
SOURCE Mus musculus (house mouse)
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
Deng, W. and Lin, H.
Miwi, a murine homolog of piwi, encodes a cytoplasmic protein
essential for spermatogenesis
Unpublished
JOURNAL
REFERENCE 2 (bases 1 to 4064)
AUTHORS Deng, W. and Lin, H.
TITLE Direct Submission
JOURNAL Submitted (22-OCT-2001) Department of Cell Biology, Duke
University, Room 412, Nanaline Duke Building, Durham, NC 27710, USA
FEATURES
location/Qualifiers
1..4064
/organism="Mus musculus"
/mol_type="mRNA"
/db_xref="taxon:10090"
/chromosome="5"
/tissue_type="testis"
191..2779
/notes="Piwi-like protein"
/codon_start=1
/product="MIWI"
/protein_id="AAL31014.1"
/db_xref="GI:16905061"
/translation="MTGRARARARGAQETVQHVAAASQOQYIPRPOOSPTGK
DIVGRGQVMVVGATSKSOELQISAGFQELSLAERGRDRPDLGVNTRQNLBHV
ESKTSGGIIVKLSTNHFRLTRSPQWALYQYHIDYNPLMEARRLSALLFQHEDLIGR
CHAFDTILFLPKRLQHKVTEVPSOTRNGEHVRIITITNELPPTSPCLQFYNIIFR
RLIKTNLQOIGRNVNPSDPIDIPNHLVWPGFTTSILOYENIMLCTDYSHKVLIR
SETVLDPMENLYQOEEHFKFOVSKELIGLVLFKYNKTYRVDIDWDQNPSTFK
KADGEVSELEYRKQINQEIIDLQPVLSQPKRRRGGPGLDIPGAMLIPBELCLITG
LTDKMRDNFNMKDLAVHTRLPQEQREVGRLLIDYIHKDDNVQRELRLDWGLSPDNL
LSFSGRIQSEKIHQGGKTFDYNPQFADWSKETRAPLISVKPDLNLLIYRNYEA
ANSLIQLFKVTPAMGIOMKAIMIEVDRTYALRLOOKVTSDIOIVVCLISNRK
DKVADLTKVLCDDPTPSOCVARTLGKQOYVMAJATKALOMCKMGSELWRVDMPL
KIAMVIGIDYHDITAGRSIAGFVASINEGTRFSPRCVDFDRGSELVDGLKVLQQA
AURANGCNEYSPSRVIVTRDVGQDLATLVNIEVPQFDLCLKSVGRGINPRLTIV
VKRVNARFFAQSGRLQPLEGTIDVEVTPWEDYFFIVSQAVRSGSVSPTHYNI
YDSSELKPDHIOQLATYKLVHYNYPGVIRVPAPCOYAHKLAFLVQSIHREPNLSLS
NRLYL"
BASE COUNT 1115 a 982 c 1077 g 890 t
ORIGIN

Query Match 72.4%; Score 1684.6; DB 10; Length 4064;
Best Local Similarity 82.8%; Pred. No. 0;
Matches 1924; Conservative 0; Mismatches 399; Indels 0; Gaps 0;

Qy 6 CTCTGTGTGAACACAGGACGACACTTAGCACCATGTATAAGATCAAAACAGGTTCTTC 65
Db 457 CTTGTGTGAACACAGGACGACACTTAGCACCATGTATAAGATCAAAACAGGTTCTTC 516
Qy 66 AGSCATTATAGTAAAGTAAAGCACTAACCATTTCCGGCTGACATCCCGTCCCGAGTGGC 125
```

Db 1597 GTTTGATTACACCCACCAATTTGCAGACTGGTCCAAAGAAACAAGAGCGCGCCGCTGAT 1656
 QY 1206 TAGTGTAAAGCCACTAGATACTGGCTGTGTATCTATACGCGAAGAAATATGAAGCAGC 1265
 Db 1657 CAGCGTGAAGCCATTGGATAAATCTGCTGCTGCTATATACCCGACGAAATATGAAGCAGC 1716
 QY 1266 CAATTCATGTATACAAAATCTATTAAAGTTACACGAGCCATGGGCATCAATGAGAAA 1325
 Db 1717 CAACCTCACTGATACAGAACTGTTCAGAGTACTCCAGCCATGGGCATCAGATGAAAA 1776
 QY 1326 AGCAATAATGATTGAAGTGGATGACAGAACTGAAGCTACTTAAGAGTCTTACAGCAAAA 1385
 Db 1777 GCATCATGATCGAGGTGGATGACAGAAACAAGCTTATCTGAGAGCCTTCGACAGAA 1836
 QY 1386 GGTACAGCAGACACCCAGATAGTTGCTGCTGTTGTCAGAAATCGGAAGGACAAAATA 1445
 Db 1837 GTGTGAGCTCAGACACTCAGATAGTTGCTGCTGTTGTCAGAAATCGGAAGGACAAAATA 1896
 QY 1446 CGATGGTATTAATAAATACCTGTGTACAGATTGCCCTACCCAGAGTCAGTGTGTGTGGC 1505
 Db 1897 TGATGCCATCAAGAAGTACTTGTGTACAGACTGCCCCACCCCAAGTCACTGTGTGTGGC 1956
 QY 1506 CCGAAGCTTAGGCAACACAGAACTGCTCATGGCCATTGCTACAAAGATTGCCCTACAGAT 1565
 Db 1957 CCGAGCCCTGGGCAAGCAAGCAAGTCAATGGCCATTGCCCAAGATGCCCTGAGAT 2016
 QY 1566 GAACTGCAAGATGGGAGGAGAGCTCTGGAGGTGGACATCCCCCTGAAGCTCGTGTATGAT 1625
 Db 2017 GAACTGCAAGATGGGAGGAGAGCTCTGGCGGGTGGACATGCCCTGAAGCTCGCAATGAT 2076
 QY 1626 CGTTGGCATCGATTGTCATGACATGACATGACAGCTGGCGGAGTCAATCGAGGATTTGT 1685
 Db 2077 CGTGGCATCGACTGTATCATGACACCAAGCTGGGGGAGGTCCATCGCAGGATTCGT 2136
 QY 1686 TGCCAGCATCAATGAAGGATGACCGCTGTTCTCAGAGCTGATATTTACAGATAGAGG 1745
 Db 2137 CGCCAGCATCAATGAAGGATGACCGCTGTTCTCCGCTGGCTTTTCAGAGCCGCGG 2196
 QY 1746 ACAGAGCTGTGTAGTGGCTCAAAGTCTGCCCTGCAAGCGGCTCTGAGGCTTGAATAG 1805
 Db 2197 GCAGAGCTGTGTAGTGGCTCAAAGTGTCTTGAAGTGTCTCTGAGGCTTGAAGTGG 2256
 QY 1806 CTGCAATGATGATCATGCCAGCGGATCATCTGTACCGGATGGCGTAGAGAGCGGCA 1865
 Db 2257 CTGCAATGATGATCATGCCAGCGGCTCATCTGTGTACCGGATGGCGTAGAGAGCGGCA 2316
 QY 1866 GCTGAAAACACTGGTGAACCTACGAAGTGCACAGTTTTTGGATTGTCTAAATCCATTGG 1925
 Db 2317 GCTGAAGACCTGGTCAATTATGAGTCCCAAGTTCCTAGATTGCTCAAGTCACTCGG 2376
 QY 1926 TAGAGTTACACCTTAGACTTAACGGTAATTGTGGTGAAGAAAGAGTGAACACCAAGATT 1985
 Db 2377 GAGAGTTACACCCCAAGACTGACTCTAATCTGGTGTGAAGAGCGGTGTAATGCCAGTT 2436
 QY 1986 TTTTGTCTAGCTGAGAGGAGACTTCAGAACTCCATTCCTGGAACAGTTATTGATGATGA 2045
 Db 2437 TTTTGTCTAGCTGAGAGGAGACTTCAGAACTCCATTCCTGGAACAGTTATTGATGATGA 2496
 QY 2046 GGTATACAGACCAAGATGGTATGACTTTTTTATCGTGAGCCAGGCTGTGAGAAAGTGGTAG 2105
 Db 2497 AGTCACAGACCAAGATGGTATGACTTTTTTATCGTGAGCCAGGCTGTGAGAAAGTGGTAG 2556
 QY 2106 TGTTCCTCCACATTCATCATGATGATGATGATGATGATGATGATGATGATGATGATGAT 2165
 Db 2557 TGTTCCTCCACATTCATCATGATGATGATGATGATGATGATGATGATGATGATGATGAT 2616
 QY 2166 ACAGCGCTTACCTACAGCTGAGCAGATCTATTACACTGCGCAGGCTGATTCGTTCT 2225
 Db 2617 CACGCGGCTGACATACAGCTTGCACCGTGTACTATAATTGGCCCTGGAGTATCCAGAT 2676
 QY 2226 TCCTGCTCTCTGAGTACGCCACCAAGCTGGCTTTTCTGTTGGCCAGAGTATTCACAG 2285
 Db 2677 CACTGACCTTGCATGATGACACAGCTGGCTTCTCTGTTGGCCAGAGTATTCACAG 2736

QY 2286 AGAGCCAAATCTGCTACTGTCAAAACCGCCTTTACTACTCTAA 2328
 Db 2737 AGAGCCAAACCTCTCCCTGTCAAAACCGCCTTTACTACTCTAA 2779

RESULT 8
 AF336369 3273 bp mRNA linear VRT 24-JAN-2003
 LOCUS
 DEFINITION
 AF336369
 ACCESSION
 AF336369.1 GI:18033486
 VERSION
 KEYWORDS
 SOURCE
 ORGANISM
 Danio rerio (zebrafish)
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Actinopterygii; Neopterygii; Teleostei; Ostariophysi;
 Cypriniformes; Cyprinidae; Danio.
 REFERENCE
 1 (bases 1 to 3273)
 AUTHORS
 Tan, C.-H., Lee, T.-C., Weeraratne, S.D., Korzh, V., Lim, T.-M. and
 Gong, Z.
 TITLE
 Zili, the zebrafish homologue of the Drosophila piwi:
 co-localization with vasa at the embryonic genital ridge and
 gonad-specific expression in the adults
 JOURNAL
 REFERENCE
 2 (bases 1 to 3273)
 AUTHORS
 Weeraratne, S.D., Gong, Z. and Tan, C.-H.
 TITLE
 Cloning and characterization of zebrafish homolog of piwi,
 essential for germ-line stem cell self-renewal
 JOURNAL
 REFERENCE
 3 (bases 1 to 3273)
 AUTHORS
 Weeraratne, S.D., Gong, Z. and Tan, C.-H.
 TITLE
 Direct Submission
 JOURNAL
 Submitted (15-JAN-2001) Department of Biological Sciences, National
 University of Singapore, 10 Kent Ridge Crescent, Singapore 119260,
 Singapore

FEATURES

Location/Qualifiers
 1..3273
 /organism="Danio rerio"
 /mol_type="mRNA"
 /db_xref="taxon:7955"
 64..2640
 /codon_start=1
 /product="piwi protein"
 /protein_id="AAL57170.1"
 /db_xref="GI:18033487"
 /translation="MTGRARSRGRGQBPAPGAQPPVQSQAAPVTPSPBQGL
 VGRGKPAKAMSEEMLQISAGFOVKIGERGRRDFHDSGVHTQLMEHVKEK
 TGVSGTALERANFWLLSRPMWALYQVVDYKPMESRLRSALLFOHEETLGKAT
 FDGAILFLPNKRNATVLCSETRNGEKVEITVLTNLPSPVCLQFYNIFRRL
 RILNMQQGRHYNPDPFNIPOHRLTIWPGFTTILQYESSIMLCSDVSHKLRSET
 VLDYMSLRQCGQRFPEACTKELVGLIILTKYNNKYRIDDIADWHTPNNTFKGD
 TEISFNKYFQSYGLDITDGNQVLLVSHVKRLGSPGPPAMLVPEFCVLTGDK
 MRADFNIMKDLASHTLSPEOREGRINRLISINRNGDVONELTTWGLSFENKLISLN
 GRVLPSELIQGRAFENPWTADSKEMRGLPLISCMSLDNLMFWYRRNADVAQSL
 LQTLKSGPMGRMQRAMVIEYEDRQESLLRALQONVARETQMVLVLPNPKDKYD
 CQKYLVCDCPTQCVSRITISKPTADMTATKIALQNMCKNGMELWSVEIPLQML
 IVGIDYHDHTAAGSRICAMVASLNQGSRFKVLQNRQGEIIDLKGLSQALKA
 YLKNYNSLPRIIVRDGVMGLQSVVDVPEQIMQIKTMQGDYEPKLSVTVVKRR
 ISSPFARIIDGKIAMPPEGTVIDTEVTRPEWYDFIVSQAVRFGCVATHYNVFDNS
 GLKPDHMQRLTYKLCHMYNNQGIIVRVPAPQYAHKLAFLVQSHKEHPNNLDDFLY
 YL"
 BASE COUNT 922 a 717 c 787 g 847 t
 ORIGIN

Query Match 43.5%; Score 1013.2; DB 5; Length 3273;
 Best Local Similarity 65.1%; Pred. No. 3.2e-262;
 Matches 1511; Conservative 0; Mismatches 808; Indels 3; Gaps 1;
 QY 7 TTTGTGTGACACAAAGGACAGAACTTAGACCATGTATAAGAAATCAAAACAGGTTCTTCA 66
 Db 322 TCTGTGTCCATCTCGTCAACTTATGGAACACGCGTGAAGAAATCAAAACAGGTTCTTCT 381

QY 67 GGCATTATAGTAAGGTTAAGCACTAACCAATTTCCGGCTGACATCCGCTCCAGTGGGCC 126
Db 382 GGGACTGCCATTGACGTGAGGCCAATTTTATGCGTCTGCTTTTACGCTCCTATGTTGGCT 441
QY 127 TTATATCAGTATACATTGACTATTAACCACTGATGGAAGCCAGAAAGACTCCGTTTCACT 186
Db 442 TTGTATCAGTATCATGTGACTTACAAGCCACCATGGAAGTCCAGACGACTGGGTTGGCT 501
QY 187 CTTCTTTTCAACAGAGATCTAATTTGGAAGTGTCTGCTTTTGTGTAAGCAAGTATTA 246
Db 502 TTGCTTTTCCACGACGAGAGACTCTGGGGAAGCACAACCTTTGTGAGGCCATTTCTC 561
QY 247 TTTTACCTTAAAGACTACAGCAAAAGGTTACTGAAGTTTTTGTAAAGCCCGGAATGGA 306
Db 562 TTTCTCCCCAACAAGCTGCGCAAGCTGAGACAGTCTCTGAGTGAAGCAGCAAGTATGA 621
QY 307 GAGGATGTGAGGATTAACGATCTTTTAAACAATGAATTTCCACCTTACATCAACAATTGT 366
Db 622 GAGAAGGTTGAGTCACTGTGACACTGACCAATGAGTCCGCCCGCTCTCTCTGTATGC 681
QY 367 TTGCAAGTTCTATAATATTATTTTCAGGAGGCTTTTGAAAATCATGAATTTGCAAAATT 426
Db 682 CTGCAAGTTTTCACAATCTCTTCAGGAGAACTCTTAAGAAATCTAAACATGACGAGATC 741
QY 427 GGCAGAAATTTATTAACCCAAATGACCCAAATTTGATTTTCCAGTCAACAGTTGGTGAT 486
Db 742 GGAAGTCAATTACTACAACTCTGATGACCCATTCAACATCCACAGCAGCAGGCTGACAA 801
QY 487 TGGCTGGCTTCACTACTTCCATTCCTTCAGTATGAAACAGAGATCATGCTCTGCACTGAC 546
Db 802 TGGCTGGCTTCAATGACCACTCTTACAGTATGAGTCCAGTCCAGCATCATGCTGTGCTGAT 861
QY 547 GTTAGCAATAAGTCTTCCAGTATGAGTCTTTTGGATTTTCATGTTCAACTTTTATCAT 606
Db 862 GTGACCAACAGGTTCTGCGAGTGAAGCGTTCTTGAATTTAATGACAGTCTGAGACAG 921
QY 607 CAGACAGAAAGACATATAATTTTCAAGAACAAAGTTTCCAAAGAACTAATAGTTTGTGTT 666
Db 922 CAGTGTGAGACAGCGCTTCCCTGAAGCTGCAAGAGAGCTTGTGGCTGTGATCAT 981
QY 667 CTTACCAAGTATAACAAATAGACATACAGAGTGGATGATTTGACTGGGACAGAAATCCC 726
Db 982 CTGACCAAGTACAACAACCAAGACATACAGAGTATGATGACATTTGCCCTGGGATCATCTCC 1041
QY 727 AAGAGACCTTTTAAGAACCCGAGGCTCTGAGTCAAGTCTTTAGAAATACATACAGGAG 786
Db 1042 AACAACACATTTCAAGAAAGGAGA---TACAGAAATTTCAATCAAGAACTACTTCAATCG 1098
QY 787 CAATACAAACCAAGAGATCACCGACTTTGAAGCAGCCTGTCTTGTGACAGCCGCCAAGAGA 846
Db 1099 CAATATGGTCTTGACATACAGATGCAACCGAGTTCTTCTGTTAGCCATGTGAAGAGA 1158
QY 847 AGCGGGGCCCTGGGGGACATGCGCAGGCGCTGCGCATGCTCATTTCTGAGCTCTGCTAT 906
Db 1159 CTTGGCCCTTCAGGACGCGCTTCCAGCGGACCAAGCTATGTTTCTGCCCGAGTTCTGCTAC 1218
QY 907 CTTACAGGCTTAACATATAAATGCGTAAATGATTTTAAACGTGATGAAGACTTTAGCCGTT 966
Db 1219 CTCACCTGGTCTAACCCGATATAATGCGTGTGATTTTAAACATGAAGGATCTTGTCTCA 1278
QY 967 CATACAGAGTAACTCCAGAGCAAGGACGCTGGAAGTGGGAGCTCAATTTGATTAATTT 1026
Db 1279 CATACAGGCTAAGTCTTGAACAGAGAGAGGCGAGAAATTAACCGCTTAATCTCCAATC 1338
QY 1027 CATAAAAACGATTAATGTTTCAAGGGAGCTTCGAGACTGGGGTTTGAAGCTTTGATTTCAAC 1086
Db 1339 AACAGAAATGGACAGCTGCAAGATGAATGACTTGGGGTCTGAGCTTTGAGAAACAAG 1398
QY 1087 TTACTGCTCTTCAGGAGAAATTTTGCAAAACAGAAAGATTCACCAAGGTGAAACAA 1146
Db 1399 CTTCTTAGTCTTAATGGGAGAGTTTGGCCCTCTGAGAGGATCATACAGGAGGACAGACA 1458
QY 1147 TTTGATTACAAATCCACAAATTTGACAGATTGGTCCAAAGAAACAGAGGTGCAACCAATTA 1206

Db 1459 TTTGAGTATAACCCATGAGCAGCAGACTGGTCTTAAAGAGATGAGAGGTCTCCCTCTGAT 1518
QY 1207 AGTGTAAAGCCACTAGATAACTGCTGTTGATCTATAGCGGAAGAAATTAAGACAGCC 1266
Db 1519 AGCTGCAATGCTCTGGAACAACCTGCTCATGTTCTACAGCGCAGAAATGCAAGATGTTGCC 1578
QY 1267 AATTCAATTGATACAAAATCTATTTAAAGTTTACACAGCCATGGGCTATCAAAATGAGAAA 1326
Db 1579 CAATCTCTGCTTCAAGCTCTCAACAAAGTGTCTGGACCCATGGGATCCGATGACAGAG 1638
QY 1327 GCAATAATGATTAAGTGAATGACAGAACTGAAGCTTCTTAAGAGTCTTACAGCAAAAG 1386
Db 1639 GCTGTGATGATTAAGTATGAAGATCGTCAGGAATCTCTCTTAGAGCTTTGCAACAAAAC 1698
QY 1387 GTCAAGCAGACACCCAGATAGTTGTCTGTCTGTCTGTCATGTAATTCGAAGGACAAATAC 1446
Db 1699 GTAGCAGCTGAACTCAGATGTTGGTGTGATCTTGGCTACCAATCGAAAGACAAAGTAC 1758
QY 1447 GATGCTATTTAAAAATACCTGTGACAGATTTGCCCTACCCCAAGTCACTGTGTGTGGCC 1506
Db 1759 GACTGTGTAAGAAGTACTCTGTGTGAGCTGCCCTACTCTCTAGTCACTGTGTGTGCT 1818
QY 1507 CGAACCTTTAGGCAACACAGCAACTGTCTATGGCCATTTGCTACAAAGATTTGCCCTACAGATG 1566
Db 1819 CGCACTATCAGCAAAACCTCAAGCACTTATGACTGTGGCCACCAAGATTTGCTTTGAGATG 1878
QY 1567 AACTGCAAGATGGGAGGAGAGCTCTGAGGCTGGACATCCCTCTGAAGCTCTGTGTATGATC 1626
Db 1879 AACTGCAAAATGGGAGGAGAGCTGTGAGTGTAGAAATTCACCTCCGAGAGCTTATGAT 1938
QY 1627 GTTGCACTCGATTTGTACCATGACATGACAGTGGGCGAGGTCAATCGAGATTTGTT 1686
Db 1939 GTAGTATAGACTGTCTACCATGATACCGCTCTGTGAAAAGGTCCATCGGCAATGTTG 1998
QY 1687 GCCAGCATCAATGAAGAGGATGACCGCTGTTTCTCACGCTGCAATTTTCAGGATAGAGA 1746
Db 1999 GCCAGCTTAAATCAGGCAATGTCAAGTGTGTTCTCCAAGTGTCTTCAAAACCGTGA 2058
QY 1747 CAGGAGCTGTAGATGGGCTCAAACTCTCCCTGCAAGCGCTCTCAGGCTTTGGAATAGC 1806
Db 2059 CAGGAGATCATCGATGCACTGAAGGGTCTATTGCAAGGTGCTCTGAAGCTCTTCTGAAA 2118
QY 1807 TGCAATAGTATACATCCCGAGCGGATCATCGTGTACCGGATGGCTTAGGACGCGCAG 1866
Db 2119 TACAACAATTTCTGCTCCCTCGGCAATCATTTTACCGAGATGGAGTTGGAGATGGCATG 2178
QY 1867 CTGAAAACACTGTGTGAATGAGAGTCCACAGTTTTTTGGATTTGTCTAAAATCCATTTGT 1926
Db 2179 CTGCAGAGCGTTGTGGAATGGAATCTCTCAGATCATGCACTCTATCAAGACCAATGGG 2238
QY 1927 AGAGGTTACACCTTAGACTAACGTAATCTGTGTGAAGAAAGAGTGAACACCAATTT 1986
Db 2239 CAAGATTTAGGCCCAAGCTCTCAGTGTGGTGGTGAAGAAAGCCATCAGCTCCCGATTC 2298
QY 1987 TTTGCTCAGTCTGGAAGGAGACTTCAGAAATCCACTTCTCTGGAACAGTTATTGATGTAG 2046
Db 2299 TTTGCTCGGATTTGATGGCAAAATTCGCAATCCCTCTCTGGGACAGTCACTGACACTGA 2358
QY 2047 GTTACAGACCAAGATGATGACTTTTATCTGTGAGCCAGCTGTGAGAAAGTGGTAGT 2106
Db 2359 GTCACTGTCCAGAGTGTGATGATTTCTTCTTATTTGAAGCCAGGCTGTGGCTTTGGGT 2418
QY 2107 GTTCTTCCACACATTAATGTCAATGTCACTATGACACAGGCGCTGAAGCCAGACACATA 2166
Db 2419 GTGCGACCTTACTACTACAAATGTGTGTTTGAACAAGCGGCTTTAAACCAATCACATG 2478
QY 2167 CAGCGCTGACCTTACAGCTGTGCACATCTATTATCACTGCCAGGTGTCTATTCTGTGTT 2226
Db 2479 CAGAGACTCACCTTACAACTTTGGCAATGTACTAACTGGCAGGGATTTGTGGAGTGT 2538
QY 2227 CTTGCTCTTTCAGTACGCCCAAGCTGGCTTTTCTTGTGTGGCCAGAGTATTCACAGA 2286

Db 2539 CCAGCTCCATGCGCATATGCCACAAATGGCTTCTCTGGTGGCCAGAGCATTCACAAG 2598

QY 2287 GAGCCAAATCTGTCACTGTCAACCCGCTTTACTACCTCTAA 2328

Db 2599 GAGCCTAACATGAATCTGGATGACTTCTCTGTATTACCTGTAA 2640

BC031060 3006 bp mRNA linear PRI 13-JUN-2002

Homo sapiens, Similar to piwi like homolog 1 (Drosophila), clone

MGC:33296 IMAGE:5271278, mRNA, complete cds.

BC031060

BC031060.1 GI:21410560

MGC.

Homo sapiens (human)

ORGANISM

Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

1 (bases 1 to 3006)

Strausberg,R.

Direct Submission

Submitted (03-JUN-2002) National Institutes of Health, Mammalian

Gene Collection (MGC), Cancer Genomics Office, National Cancer

Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590,

USA

NIH-MGC Project URL: <http://mgc.nci.nih.gov>

Contact: MGC help desk

Email: cgabbs@mail.nih.gov

Tissue Procurement: Miklos Palkovits, M.D., Ph.D.

CDNA Library Preparation: Michael J. Brownstein (NHGRI) & Shiraki

Toshiyuki and Piero Carninci (RIKEN)

CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)

DNA Sequencing by: Sequencing Group at the Stanford Human Genome

Center, Stanford University School of Medicine, Stanford, CA 94305

Web site: <http://www.shgc.stanford.edu>

Contact: (Dickson, Mark) mcd@paxil.stanford.edu

Dickson, M., Schmutz, J., Grimwood, J., Rodriguez, A., and Myers,

R. M.

Clone distribution: MGC clone distribution information can be found

through the I.M.A.G.E. Consortium/LLNL at: <http://image.llnl.gov>

Series: IRAK Plate: 48 Row: b Column: 15

This clone was selected for full length sequencing because it

passed the following selection criteria: Similarity but not

identity to protein.

Location/Qualifiers

1. .3006

/organism="Homo sapiens"

/mol_type="mRNA"

/db_xref="taxon:9606"

/clone="MGC:33296 IMAGE:5271278"

/tissue_type="Testis"

/clone_lib="NIH MGC 97"

/lab_host="DH10B"

/note="Vector: pBluescript"

4. .2562

/codon_start=1

/product="Similar to piwi like homolog 1 (Drosophila)"

/protein_id="AAH31060.1"

/db_xref="GI:21410561"

/translation="MSGARVKGARGIARSPSATEVGRIOASPLPSVDLSNNEASSSN

GLGTSRTNINIKYIGSSGDAGSTFMERGVKNQDFMDLSICTREKIAHVRNCKTGSS

GIPVLVTLNFDLPDDWOLYOYHVYIPDLASRLRLIALLYSHSELSNKAQFDGA

ILFLSOKLEKVTESSETGETTGMTITLRELPSFVCIQVFNIIIFRILKLS

MTQIGRNFYNSPEMIEIPOHKLMLPFGPAISYFPERKLFLFADSVYKVLNRTVLEF

MTALCORTGLSCFTCEKQOLJGLIVLTRYNNRTYSIDIDWSVKPHTFOKRDTEI

TVVDYKQOYDITVSDNLPMPLYSLLKERNNSAQALHPLCELTGLTGLDQATSD

FQMLKVAZKTLSPSGROQLRLVDNIQRNINARFELTWGLHFGSISLTGRIVP

SEKILMQDHIQFVSAADWSKDIRTCILNLAQSLNTWLILCSDRTEYVAESFLNCLR

VAGSMGVNDYIPKIKVENPAAFRAIQOYVDPDVLVLCIPSNQKTYDSIKKLL

SSDCPVPSOCVLARTLNKGMMSIATKIAMQMTCKLGGELWALBIPKLSMVGVIDV

CKDALSQVWVGCVASVNPRTFRFSCILQRTMTDADCLKVFMTGALNKWYKYNH

Query Match 29.2%; Score 680; DB 9; Length 3006;

Best Local Similarity 56.7%; Pred. No. 3.7e-172;

Matches 1319; Conservative 0; Mismatches 1000; Indels 9; Gaps 3;

BASE COUNT 904 a 627 c 674 g 801 t

ORIGIN

1 ATGATCTTTGTGTGTAACACAGGAGCAATAGACCAATGTAAGAATCAAAAACAGT 60

244 ATGATTTGATGATCTGTACACAGAAATGTCACATGTGAGAAATGTAAGAAGT 303

61 TCTTCAGGCATTTAGTAAGGTTAAGCACTAACCAATTTCCGGCTGACATCCCGTCCCCAG 120

304 TCCAGTGAAATACCTGTGAAACTGGTTACAAACCTCTTTAACTTAGATTTTCCCCAAGAC 363

121 TGGGCTTTATATCAGTATCAGATTCATTAACCCACTGATGAAGCAGCAGAGATCCGT 180

364 TGGCAGCTATACCACTACCATGTGACATATATTCAGATTTTAGCATCTAGAAGGCTGAGA 423

181 TCAGCTCTCTTTTCAACACAGAGATCTAATTTGAAAGTGTCTGCTTTTGTATGAGCAAG 240

424 ATTGCTTTTACCTTTAGTCAATGAGACTTTCCACAAAGCAAGCAATTCGACGGTCC 483

241 ATATTTATTTTACCTTAAAGACTACAGCAAAAGGTTACTGAAGTTTTTAGTAAGACCCGG 300

484 ATCTCTTTTCTGTACAAAGCTAGAGAAAGGTCACAGAGTTGTCAAGTGAATCAAACTCAA 543

301 AATGAGAGGATGTGAGGATAACGATCTTAAACAATGAATCTCCACCTACATCACCA 360

544 AGAGTGAGACTATAAAGATGACTATCACCTGAAGAGGGAGCTGCCATCAAGTTCTCCC 603

361 ACTTGTGTGAGTTCTATAATATTATTTTTCAGAGGCTTTTGAATCATGAATTTGCAA 420

604 GTGTCATCCAGGCTTCAATATCATCTTCAGAAAGATCCTCAAAAAGTTGTCCATGTAC 663

421 CAATTTGACGAAATTTATTAACCAATGACCAATTTGATTTTCCAAAGTCACAGGTTG 480

664 CAATTTGACGAACTTCTATAATCTTTCAGAGCAATGGAATTTCCCCAGCACAAATTA 723

481 GTGATTTGGCTTGGCTTCACTACTTCCATCTCTAGTATGAAGACAGCATCATGTCTGC 540

724 TCCCTTTTGGCTGGGTTTGGCAATTTCTGTGCATATTTTGAAGGAAGCTCTCTGTAGT 783

541 ACTGACGTTAGCCATAAAGTCTTCGAAGTCAGACTGTTTGGATTTTCATGTTCAACTTT 600

784 GCTGATGTGAGTTACAAAGTCTTCGGGAATGAGAGGTTCTTGGAAATTCATGACTCTC 843

601 TATCATCAGACAGAAACATAAATTTCAAGAACAAAGTTTCCAAAGACTAATAGGTTTA 660

844 TGTCAAGAACTGGCTTCTCTCTGTTTCCACCAGAGCTGTGAGAGCAGCTAATAGGGCTC 903

661 GTTGTCTTCAAGTATAACAATAAGACATACAGAGTGGATGATTTGACTGGGACCCAG 720

904 ATTGCTCTTCAAGATACATAAGACCACTTCTCAATGTAGTATGATTTGCTGGTCAGTG 963

721 AATCCAAAGACACCTTTTAAAGAACCCGAGCGCTCTGAAGTCAGCTTCTTAGAATACTAC 780

964 AAGCCACACACACCTTTTCAGAGCGGATGGCACCAGATCACCTATGTGATTTACTTAC 1023

781 AGGAAGCAATCAACCAAGAGATCACCGACTTGAAGCAGCCTGTCTTGGTCAGCCAGCCC 840

1024 AAGCAGCAGTATGATTTACTGTATCGACCTGAATCAGCCCATGCTTGTGTAGTCTGTTA 1083

841 AAGAGAGCGCGGCGCTGGGGGAGCACTGCCAGCGCTGCCATGCTCATTTCTCTGAGCTC 900

1084 AAG---AAGAGAGAAATGACAAAGTGAAGGCTCAGCTCGGCCACCTGATACCTGAGCTC 1140

901 TGCTATCTTACAGGTTAACTGATAAATGCGTATGATTTTAAAGTATGATAAAGACTTA 960

1141 TGCTTTTAAAGGCTGACTTGACCAAGCAACATCTGATTTCCAGCTGATGAAGCTGTG 1200

QY	961	GCCTTTCATACAAGACTAACTCCAGAGCAAAAGCAGCGTGAAGTGGGACGACCTCATTTGAT	1020
Db	1201	GCTGAAAAGACACGCTCTCAGTCTCTCAGGCGCGCAGCAGCGCTGTGGAC	1260
QY	1021	TACATTCATAAAACAGTAATGTTCAAAGGAGCTTCGAGACTGGGGTTTGACCTTTGAT	1080
Db	1261	AACATCCAGAGGAATACCAATGCTCGCTTTGAACTAGAGACCTGGGAGCTGCATTTTGA	1320
QY	1081	TCCAACTTACTGTCTCTCAGGAAGAAATTTGCAACACAGAAAAGATTACCAAGGTGGA	1140
Db	1321	AGCCAGATA--TCTCTGACTGGCCGGAATTTGGCTTCAGAAAAATATTAAATGCAAGAC	1377
QY	1141	AAAACAATTTGATPACAAATCCACAATTTGCAGATTTGGTCCAAAGAAAACAAGAGGTGACCA	1200
Db	1378	CACATAT---GTCAACCTGTGTCTGCTGCTGACTGGTCCAAAGGATATTCCAATCTGCAAG	1434
QY	1201	TTAATTAGTCTTAAGCCACTAGATACTGCTGTTGTATCTATACGCGAGAANTATGAA	1260
Db	1435	ATTTTAATGCAAGTCTTTGAATACCTGGTTGATTTTATGTAGCAGACACTGAATAT	1494
QY	1261	GCAGCCAAATTCATTGATACAAAATCTATTAAAGTTACACAGCCATGGGCATGCAAAATG	1320
Db	1495	GTTCGCCAGAGCTTTCTGAACCTGCTCAGAAAGAGTTGCAGGTTCCATGGGATTTAATGTG	1554
QY	1321	AGAAAGCAATATGATTTGAGTGGATGACAGAACTGAAGCCCTACTTTAAGAGTCTTACAG	1380
Db	1555	GACTACCCCAAAATCATAAAAGTACAGAAAATCCAGCTGCATTTGTTAGAGCTATACAG	1614
QY	1381	CAAAAGGTTCAGCAGACACCCAGATAGTTGTCTGCTGTGTTCAAGTAATCGGAAGGAC	1440
Db	1615	CAATAGTTGATCTTGATGTTTCACTGGTAAATGTGCAATTCGCTTCTAATCAGAGACC	1674
QY	1441	AAATACGATGCTATTAAAAAATACCTGTGACAGATTCCTTACCCCAAGTCAAGTGTGTG	1500
Db	1675	TATTTATGATTCCTATTAAAAAATATTTTCAGCTCAGACTGCCAGTCCCAAGCCAAATGTGTG	1734
QY	1501	GTGGCCGACCTTAGCCAAACAGCAACTGTCTATGSCCAATGCTCAAGAAGTTGCCCTA	1560
Db	1735	CTTGCTCGGACCTTGAAATAACAGGGCATGATGATGATAGTATCGCCACCAAGATCGGTATG	1794
QY	1561	CAGATGAATCGCAAGATGGGAGGAGAGCTCTGGAGGGTGGACATCCCCCTGAAAGCTCGTG	1620
Db	1795	CAGATGACTTGCAGCTCGGAGCGAGCTGTGGCTGTGGAAATACCTTTAAAGTCCCTG	1854
QY	1621	ATGATCGTTGGCATCGAATGTTTACCATGACATGACAGCTGGGCGAGGTCAATCGCAGGA	1680
Db	1855	ATGGTGTGCGTATTGTATGTCTGTAAAGATGCACTCAGCAAGGACGTGATGGTGTGGGA	1914
QY	1681	TTTGTTCCGAGCATCAATGAAGGATGACCGCTGGTTCTTCAGCTGTCATATTTTCAGGAT	1740
Db	1915	TGCGTGCAAGTGTAAACCCAGAAATCACAGGTGGTTTTCCGCTGTATCTCCTTCAGAGA	1974
QY	1741	AGAGGACAGGAGCTGGTAGATGGCTCAAAAGTCTGCCTGCAAGCGGCTCTGAGGGGCTTGG	1800
Db	1975	ACAATGACTGATGTTGCAGATTTGCTTGAAGTTTTCATGACTGGAGCACTCAACAATGG	2034
QY	1801	AATAGCTGCATAGTACATGCCAGCGGGAATCATCGTGTACCGCATGCGGTAGAGAGAC	1860
Db	2035	TACAAGTACATCATGATTTTGCACAGCGGATAATTGTGTACCGTGTGCTGTAGGGGAT	2094
QY	1861	GGCCAGCTGAAAACAACCTGGTGAACTACGAGTGCACAGTTTTTGGATTCTCTAAATCC	1920
Db	2095	GGTCAGCTGAAAACAACCTATTGGAATATGAAGTCCACAGCTGTGACGACTGTGGCAGAA	2154
QY	1921	ATTGGTAGAGGTTACAACCCCTAGACTAACCGTAATTTGTGTGAAGAAAAGAGTGAACACC	1980
Db	2155	TCCAGCTCAATAACACGCTCAAGACTGTGCGTGTATTTGTGTTCAGGAAGAAGTGCATGCCA	2214
QY	1981	AGATTTTTTGTCTAGCTGAGGAGGAAGCTTCAGAAATCCAATTCCTCTCGAAACAGTTATTGAT	2040
Db	2215	CGATTTCTTACCGAAATGAACCGCACCTGTACAGAACCCCACTTTGCACTGTTGTGGAT	2274

QY	2041	GTGAGGTTACCAAGACCAAGATGGTATGACTTTTTTATCGTGAGCCAGCGCTGTGAAAGT	2110
Db	2275	TCAGAAGCAACACGTAACGAATGGTATGACTTTTATCTGATCAGCCAGCTGGCGCTGCCGG	2334
QY	2101	GGTAGTGTTCCTCCACACATTTACATATGTCATCTATGACAAAGAGCGGCTGAAGCCAGAC	2160
Db	2335	GGAACTGTGTAGTCTCTACCTACCTATATAATGTCATCTATGATGACAAAGGCTTGAAGCCGAC	2394
QY	2161	CACATACAGCGCTTGACTACAGCTGTGCCACATCTATTCAACTGGCCAGGTGTCAAT	2220
Db	2395	CATATCGAAGACTTACATTCAAATTTGGCCCTGTACTACAATGGCGGCGCATGTC	2454
QY	2221	CGTGTTCCTGCTCTTCGCCAGTAGTACGCCCAAGAGCTGGCTTTTCTTTGTCGCCAGAGTATT	2280
Db	2455	AGTGTCCAGCAGCAATGTGATGATGCTTCACAAAGCTGACCTTTCTGGTGGCACAAAGCATT	2514
QY	2281	CACAGAGAGCCAAATCTGTCACTGTCAAACCGGCTTTTACTACCTCTAA	2328
Db	2515	CATAAGAACCCAGCTGTGGAATTTAGCCAAACCATCTCTTCTACCTGTGA	2562
RESULT 10			
AK093475			
LOCUS	AK093475	3138 bp	mRNA linear PRI 15-JUL-2002
DEFINITION	Homo sapiens cDNA FLJ36156 fis, clone TESTI2025546, weakly similar to Mus musculus mRNA for MIWI (piwi).		
ACCESSION	AK093475		
VERSION	AK093475.1	GI:217523360	
KEYWORDS	oligo capping; fis (full insert sequence).		
SOURCE	Homo sapiens		
ORGANISM	Homo sapiens		
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.		
AUTHORS	1 Tshibashi, T., Kanehori, K., Yosida, M., Watanabe, S., Ishida, S., Ono, Y., Hotta, T., Hiraoka, S., Murakawa, K., Takiguchi, S., Kusano, J., Watanabe, M., Fujimori, K., Tanai, H., Ishida, M., Yamashita, H., Chiba, Y., Sugiyama, T., Irie, R., Otsuki, T., Sato, H., Wakamatsu, A., Ishii, S., Yamamoto, J., Isono, Y., Kawai-Hio, Y., Saito, K., Nishikawa, T., Kimura, K., Matsuo, K., Nakamura, Y., Sekine, M., Kikuchi, H., Kanda, K., Wagatsuma, M., Takahashi-Fujii, A., Oshima, A., Sugiyama, A., Kawakami, B., Suzuki, Y., Sugano, S., Nagahara, K., Masuno, Y., Nagai, K. and Isogai, T.		
TITLE	NEDO human cDNA sequencing project		
JOURNAL	Unpublished		
AUTHORS	2 (bases 1 to 3138)		
TITLE	Isogai, T. and Yamamoto, J.		
JOURNAL	Direct Submission		
COMMENT	Submitted (04-JUL-2002) Takao Isogai, FLJ Project (HRI Team); 2-6-7 Kazusa-Kamatari, Kisarazu, Chiba 292-0812, Japan (E-mail:genomics@hri.co.jp, Tel:81-438-52-3975, Fax:81-438-52-3986) NEDO human cDNA sequencing project supported by Ministry of Economy, Trade and Industry of Japan; cDNA full insert sequencing: Research Association for Biotechnology (RAB); cDNA library construction: Helix Research Institute (HRI) (supported by Japan Key Technology Center etc.); 5'- & 3'-end one pass sequencing: RAB, HRI, and Biotechnology Center, National Institute of Technology and Evaluation; clone selection for full insert sequencing: HRI and RAB; annotation: HRI and RAB.		
FEATURES	Location/Qualifiers		
source	1. .3138		
	/organism="Homo sapiens"		
	/mol_type="mRNA"		
	/db_xref="taxon:9606"		
	/clone="TESTI2025546"		
	/tissue_type="testis"		
	/clone_lib="TESTI2"		
	/note="cloning vector: pME18SFL3"		
CDS	152. .2710		
	/note="unnamed protein product"		
	/codon_start=1		
	/protein_id="BAC04179.1"		
	/db_xref="GI:21752361"		

/translation="MSGRARVVKARGIARSPSATEVGRIOASPLPRSVLNSNEASSN
 GFLGTSRSTNDKYGISSDAGSTFMERGVKQDFMDLSICTREKLAHVRNCKTSS
 GIPVKLNTLFLNDFPDQWOLQYHVTYIPDLASRLRLIALLYSHSELNKAAPDGA
 ILFLSKLEKTELSESTETKITLKLRLPSPPSVICIOVFNIIFRKLKLS
 MYOIGRNFYNPSEPMELPOHKLSPGPAISVPERKLPSADYSYKVLNRETVLEF
 MTAICQRTGLSFTCEKQKOLIGLVLIRYNNRTYSIDDDWSVAPTHFQKQRTETI
 TVDYKYQDITVSDNLNPLVSLKKRNDNSQAHLPLPELFCFTGLTDQATSD
 FOLMAKAEKTLSPSGRQRLARLDVNIQRNARFELETWGLHFGSQISLITGRIVP
 SEKILMQDHICOPVSAADKDIRTCILNAQSLNLTWILICSDRTEYVAESFLNCLR
 VAGSGFVNDYPKILIKQBNPAFAIRAOYVDVQVLMVCLIPSNOKTYDYSIKKYL
 SSDCPVPSOCVLARTLKNQMMSLATKIAMOMTCKLGLWAVSLPLKSLMVGVGDV
 CXDALSXQVNVGVCVASVNPRIIRPESRILQRTWTDVADCLKVPMWGLANKWYKXN
 DEPARLIVRAGVGQDLLEIYEVQULSSVABSGNSTSSRLSVIVVRKCMRPRFF
 TENNRIVQNPGLIGTATBNWDFYLLISQVACRGTVSPTYNVIYDNDGLKPDH
 MORLTFKLCHLYNWPFGIVSPAPQYAHKLFLVAQSIHKPSPLELANHFLY"
 BASE COUNT 912 a 678 c 714 g 834 t
 ORIGIN

Query Match 29.2%; Score 680; DB 9; Length 3138;
 Best Local Similarity 56.7%; Pred. No. 3.7e-172;
 Matches 1319; Conservative 0; Mismatches 1000; Indels 9; Gaps 3;

QY	1	ATGATCTTTGGTGTGAACAACAGGAGAACTTAGACCATGTTAAAGAAATCAAAAACAGGT	60
DB	392	ATGGATTTGAGTATCTGTACACAGAAAATTGGCACATGTGAGAAATTGTAACACAGGT	451
QY	61	TCTTCAGGCATTTATAGTAAGGTTAAGCACTAACCATTTCCGGGTGACATCCCGTCCCCAG	120
DB	452	TCCAGTGAATACCTGTGAACTGGTTACAACTCTTTAACTTAGATTTTCCCCAAGAC	511
QY	121	TGGCCCTTATACGATACATGATGACTATACCCACTGATGGAAGCAGAGACTCCGT	180
DB	512	TGGAGCTATACAGTACCATGTGACATATATTCAGATTTAGCATCTAGAGGGTGTGA	571
QY	181	TCAGCTCTCTTTTCAACAGGAGACTAAATTTGGAAAGTGATGCTTTTGTGGAACG	240
DB	572	ATGCTTTTACTTTATAGTCAATGTAACCTTCCACGAAGCAAAAGCATTCGACGTGCC	631
QY	241	ATATTTATTTTACCTAAAAGACTACAGCAAAAGGTTACTGAAAGTTTATAGTAAGCCGG	300
DB	632	ATCTTTTCTGTCACAAAGCTTAGAAGAAAGGTCACAGAGTTGTCAAGTGAATCAA	691
QY	301	AATGAGAGAGTGTGAGGATACGATCACTTTTAAACAAATGAATCCCACTCATCACCA	360
DB	692	AGAGGTGAGACTATAAGATGACTATCACCTGAGAGGGAGGTGACATCAAGTTCTCCC	751
QY	361	ACTGTGTTTCAGTTCTATATATTTTTCAGGAGGCTTTTGAAGATCATGATTTGCAA	420
DB	752	GTGTGATCCAGGCTTCAATATCATCTTCAGAAAGATCTCAAAAGTTGTCCATGTAC	811
QY	421	CAAAATTGACGAAATTTATTAACCCAAATGACCCAAATTTGAATTTCCAAAGTCAAGGTTG	480
DB	812	CAAAATTGACGAAATTTCTATATCTTCAGAGCCAAATGGAATTTCCCCAGCACAAATTA	871
QY	481	GTGATTTGCGCTGGCTTCACTACTTCCATCTTCAGTATGAAAACAGCATCATGCTCTGC	540
DB	872	TCCCTTTGCGCTGGGTTTGCCTATTTCTGTGCATATTTTGAAGGAAGCTCTCTGTAGT	931
QY	541	ACTGAGCTTTAGCCATAAAGTCTCTCGAAGTGAGACTGTTTGGATTTCAATGTTCAACTTT	600
DB	932	GCTGATGTGAGTTACAAAGTCTCCGGAATGAGACGGTTCTGGAATTCATGCTGCTCTC	991
QY	601	TATCATCAGACAGAGAAACATAAATTTCAAGAACAAAGTTTCCAAAGAACTAATAGGTTTA	660
DB	992	TGTCAAAGAACTGGCTTGTCTCTTTTCAACCCAGACGTGTGAGAAAGCAGCTAATAGGGTCT	1051
QY	661	GTGTGTTCTTACCAAGTATACAAATAGACATACAGAGTGTGATATTCAGCTGGGACAG	720
DB	1052	ATTGCTTTTACAGATACAAATACAGAACTTACTCCATTTGATGACATTCAGCTGGTCAGTG	1111
QY	721	AATCCCAAGAGCACCTTTTAAGAAAGCCGAGCGGTCTGAAAGTCAAGCTTCTTAGAATACTAC	780
DB	1112	ATGCCACACACACACCTTTTCAAGACGGGATGGCACCGAGATCACCTATGTGATTACTACT	1171

QY	781	AGGAAGCAATCAACCAAGAGATCACCGACTTGAAGCAGCCCTGTCTTGGTCAGCCAGCCC	840
DB	1172	AAGCAGCAGTATGATTAATTTACTGTATCGGACCTGAATCAGCCCATGCTTGTAGTCTGTTA	1231
QY	841	AAGAGAAGCGGGGCCCTTGGGGGAGACCTGCCAGGGCCCTGCCATGCTCATTTCTCAGCTC	900
DB	1232	AAG---AAGAAGAGAAATAGCAACAGTGAAGCTCAGCTCGGCCACTGATACCTGAGCTC	1288
QY	901	TGCTATCTTACAGGTTCAACTGATAAATTCGTAATGATTTTAACTGATGAAGACACTTA	960
DB	1289	TGCTTTTAAAGGGCTGACTGACCGGCAACATCTGATTTCCAGCTGATGAAGGCTGTG	1348
QY	961	CGCGTTTCATCAAGACTAACTCCAGAGCAAAAGGCGACCTGAAGTGGGACGACTCAATTGAT	1020
DB	1349	GCTGAAAAGACACGTTCTCAGTCCCTTCAGGCCGCGCAGCAGCGCTTGCAGGCTTGTGAC	1408
QY	1021	TACATTCATAAARACGATTAATGTTCAAGGGAGCTTCGAGACTGGGTTTGAAGCTTTGAT	1080
DB	1409	AACATCCAGAGAAATACCAATGCTCGCTTGAACCTAGAGACTCGGGGACTGCATTTTGA	1468
QY	1081	TCCAACTTTACTGTCTCTCTCAGGAAGAATTTTGCAACAGAAAGATTCACCAAGGTGA	1140
DB	1469	AGCCAGATA---TCTCTGACTGGCCGGATTGTCCCTTCAGAAAANAATTAATTCGAAGAC	1525
QY	1141	AAAACATTTGATTTACAATCCCAATTTTCAGATTTGGTCCAAAGAAACAGAGGTGCACCA	1200
DB	1526	CACATAT--GTCAACCTGTGTCTGCTGCTGACTGGTCCAAGGATATTTCGAATTTGCAAG	1582
QY	1201	TTAATGATTTTAAAGCCTAGATAACTTGGCTGTGATCTATACCGGAAGAAATTAATGMA	1260
DB	1583	ATTTTAAATGCACAGTCTTTGAATACCTGGTTGATTTTATGTAGCGACAGAACTGAATAT	1642
QY	1261	GCAGCAATTCATTTGATACAAATCTATTTTAAAGTTACACGAGCATCGGCATGCAATG	1320
DB	1643	GTGCCGAGAGCTTTCTGAACTGTCTTGAGAGAGTTGAGGTTCCANUGGATTTAATGTG	1702
QY	1321	AGAAAGCAATTAATGATTAAGTGGATGACAGAACTGAAGCCTACTTAAGAGTCTTTACAG	1380
DB	1703	GACTACCCCAAAATCATAAAGTACAGAAATCCAGCTGCATTTGTTAGAGCTATACAG	1762
QY	1381	CAAAAGGTCAAGCAGACACCCAGATAGTTGTCTGTCTGTGTTGTCAGTAATCGGAGAGAC	1440
DB	1763	CAATATGTTGATCTCTGATGTTTTCAGCTGTGTAATGTGCATTTCTGCCTTTCTAATCAGAAGACC	1822
QY	1441	AAATACGATGCTATTAAAAAATACCTGTGTACAGATTCCTTACCCCAAGTCAGTGTGTG	1500
DB	1823	TATATGATTTCCATTAATAAANAATTTGAGCTCAGACTGCCAGTCCCAAGGCCAATGTGTG	1882
QY	1501	GTGGCCCGAACTTTAGGCAACAGCAAACTGTCTATGGCCATTTGTCTACAAAGATTCGCCCTA	1560
DB	1883	CTTGCTCGGACCTTGAATAAACAAGGGCATGATGATGATGATGATGATGATGATGATGATG	1942
QY	1561	CAGATGACTCCAGATCGGGAGGAGCTCTGGAGGGTGGACATCCCTCCCTGAGCTCGTG	1620
DB	1943	CAGATGACTTGAAGCTCGGAGGGAGCTGTGGGCTGTGGAATACCTTTTAAAGTCCCTG	2002
QY	1621	ATGATCGTTGGCATCGATTGTTTACATGACATGACAGCTGGGGCGGAGGTCAATTCAGGA	1680
DB	2003	ATGTTGGTTCGTTATGATGTTCTGTAAAGATGCATCTCAGCAAGGAGCTGATGTTGTTGGA	2062
QY	1681	TTTGTTCAGCATCAATGAAGGGATGACCCGCTGGTTTCTCAGCTGTCATATTTTCAGAT	1740
DB	2063	TGCGTGGCCAGTGTAAACCCAGAAATCACCAGGTGGTTTTCCTGCTGATCTCTCAGAGA	2122
QY	1741	AGAGGACAGGAGCTGTGATGGGCTCAAGCTTCCTGCTGCAAGCGGCTCTGAGGCTTGG	1800
DB	2123	ACAAATGACTGATTTGACAGATTGCTTGAAGTTTTCATGACTGGAGCACTCAACAAATGG	2182
QY	1801	AATAGCTCAATGATACATGCCAGCGGATCATCGTGTACCGGATGGCGGTAGGAGAC	1860
DB	2183	TACAGTACAAATCATGATTTGCCAGCAGCGGATTAATTTGTTACCGTCTGGTGTAGGGAT	2242

```
QY 1861 GCCAGCTGAAACACACTGGTGAACCTACGAAGTGCACACAGCTTTTGGATTGTCTAAATCC 1920
Db 2243 GGTAGCTGGAACACACTTATTGAATATGAAGTCCACAGCTCTGAGCAGTGTGGCAGAA 2302
QY 1921 ATTGTAGAGGTTTACAACCTAGACTAAGCGTAATTTGGTGAAGAAAGAGTGAACACC 1980
Db 2303 TCCGGCTCAAAATACCAGCTCAAGACTGTCGGTGAATTTGGTTCAGGAAGAGTGCATGCCA 2362
QY 1981 AGATTTTGTCTGCTGGAGGAAGACTTCAAGATCCACTTCTCGGAACAGTTATTGAT 2040
Db 2363 CGATTCTTTACCGAAATGAACCGCACTGTACAGAAGCCGCCACTTGGCACTCTTGTGGAT 2422
QY 2041 GTAGAGGTTTACCAGACCAAGATGGTATGACTTTTTTATCGTCAGCCAGGCTGTGAGAAGT 2100
Db 2423 TCAGAAGCAACACGTAACGAATGGTATGACTTTTATCTGATCAGCCAGTGTGCGCGG 2482
QY 2101 GGTAGTGTCTTCCACACATTTACATGTCATCTATGACAAACAGCGGCTGTAAGCCAGAC 2160
Db 2483 GGAACGTGTTAGTCTTACCTACTATATAATGTCATCTATGATGACAAACGGCTTGAAGCCGAC 2542
QY 2161 CACATACAGCGCTTGACCTACAGCTGTCGCCATCTATTACAACTGCCAGTGTCAAT 2220
Db 2543 CATATGCAGAGACTTACATTCAAATGTGCGCACTGTACTCAACCTGCGCGGCGCATGTC 2602
QY 2221 CGTGTCTCTGCTTCTGCGAGTACGCCCAAGAGCTGGCTTTTCTTGTGGCCAGAGTATT 2280
Db 2603 AGTGTCGCCAGCACCAGTGTGAGTATGCTCACAAGCTGACCTTTCTGTGGCACAAGCAATT 2662
QY 2281 CACAGAGACCAAACTGTGCTACTGTCAACCGCCCTTACTACCTCTAA 2328
Db 2663 CATAAAGAACCCAGCTGTGAATTAGCCACCACTCTCTTCTACCTGTGA 2710

RESULT 11
AK096837
LOCUS AK096837 2456 bp mRNA linear PRI 15-JUL-2002
DEFINITION Homo sapiens cDNA FLJ39518 fis, clone PUAE2001188, moderately similar to Mus musculus mRNA for MIWI (piwi).
ACCESSION AK096837
VERSION AK096837.1 GI:21756416
KEYWORDS oligo capping; fls (full insert sequence).
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 Tanigami, A., Fujiwara, T., Shibahara, T., Goto, Y., Hirao, M., Shimizu, F., Wakebe, H., Ono, T., Hishigaki, H., Watanabe, T., Ozaki, K., Sugiyama, T., Irie, R., Otsuki, T., Sato, H., Wakamatsu, A., Ishii, S., Yamamoto, J., Isono, Y., Kawai-Hio, Y., Saito, K., Nishikawa, T., Kimura, K., Yamashita, H., Matsuo, K., Nakamura, Y., Sekine, M., Kikuchi, H., Kanda, K., Wagatsuna, M., Murakawa, K., Kanehori, K., Takahashi-Fujii, A., Oshima, A., Sugiyama, A., Kawakami, B., Suzuki, Y., Sugano, S., Nagahara, K., Masuho, Y., Nagai, K. and Isogai, T.
NEO human cDNA sequencing project
Unpublished
2 (bases 1 to 2456)
Isogai, T. and Yamamoto, J.
Direct Submission
JOURNAL Submitted (04-JUL-2002) Takao Isogai, FLJ Project (HRI Team) ; 2-6-7 Kazusa-Kamatari, Kisarazu, Chiba 292-0812, Japan
(E-mail: genomics@hri.co.jp, Tel: 81-438-52-3975, Fax: 81-438-52-3986)
COMMENT NEDO human cDNA sequencing project supported by Ministry of Economy, Trade and Industry of Japan; cDNA full insert sequencing: Research Association for Biotechnology (RAB); cDNA library construction: Helix Research Institute (HRI) (supported by Japan Key Technology Center etc.); 5'- & 3'-end one pass sequencing: RAB, HRI, and Biotechnology Center, National Institute of Technology and Evaluation; clone selection for full insert sequencing: HRI and RAB; annotation: HRI and RAB.
Location/Qualifiers
1..2456
/organism="Homo sapiens"
```

```

/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="PUAE2001188"
/cell_type="pulmonary artery endothelial cells (HPAEC)"
/clone_lib="PUAE2"
/note="Cloning vector: pME18SPL3-Primary culture, pulmonary artery endothelial cells"
28..2028
/note="unnamed protein product"
/codon_start=1
/protein_id="BAC04873.1"
/db_xref="GI:21756417"
/translaton="MTITLKEPLSSPVCIQVENILPRKILKLSMYQIGRNFVNPSPBMEIPQKLSLWFGFAISVYPERKLLFSADVSKYLNRNTEVLEFMTALCORTGLSCFTSCQKLGILVLTIRNNRTYSIDIDMSVKPTHTPLKRDGTEITTVDIYKQYDI TVSDNQLPMLVLLTKRRDNSEPOLAHLIPELCLTGLTDOATSDFOKMAVAKTR PVSAGQORLARLVNORNTARFLELTWGLHFGSGISLTCRIVPSKILMQHICQ LPSAASDKIRCTCKILNALSNTWLILCSDREYVAESFLNCRVLRVSGMENVDP KIIKVENPAAAFVRAIOQYVDPQVQVACILPNSKNKYVDSIKVLSDDCPVBSQVL ARTLNKGMMMSIATKIAMQNTKLGELWAVEIPLSKLVVGVGDVCKDALSKDVMV GCVASVPRITRWFSCILQRTMTDADCLKVFMTGALNKWKYNHNDLUPARIIVYRAG VGDQLKTLIEYEPQLLSVAESSNTSSLSLVIVRKKCMRPFPTFEMNRTVQNPL GTVDSATARNWYDFYLSQVACRGTSPYTNVIVDNGKLPDHPMORLTFKLCPLY YNWPGIVSPAPCOYAHKLTFLVAQSIHKESLELANHLFYL"
BASE COUNT 718 a 519 c 550 g 669 t
ORIGIN
Query Match 26.5%; Score 616.8; DB 9; Length 2456;
Best Local Similarity 57.5%; Pred. No. 4.4e-155;
Matches 1170; Conservative 0; Mismatches 857; Indels 9; Gaps 3;
QY 293 AGACCGGAATGGAGAGGATGTGAGGATAAGATCAATCACTTTAACAATGAATCCACTCA 352
Db 2 AACTCAAGAGGTGAGACTATAAAGATGATATCACCCTGAAGAGGAGTGCCTCA 61
QY 353 CATCACCACAACTGTTTGCAGTCTTATAATATATTTTTCAGGAGGCTTTTGAATAATCATGA 412
Db 62 GTTCTCCGTGTGCATCCAGTCTTCAATATCATCTTCAGAAAGATCTCAAAAAGTTGT 121
QY 413 ATTTGCAACAATTTGGAGCAAAATTTATATACCAAAATGACCAATTTGATATTCAGATC 472
Db 122 CCATGTACCAAAATTTGGACGGAATCTTATAATCTCTTCAGAGCAATGGAATTTCCCGAGC 181
QY 473 ACAGGTTGGTCAATTTGGCTGGCTTCACTACTTCCATCTTCAGTATGAAACACAGCATCA 532
Db 182 ACAAATTATCCCTTTGGCTGGGTTCGCCAATTTCTGTGTCTATATTTTGAAGGAGCTCC 241
QY 533 TGCTCTGCACCTGACGCTTAGCCATAAAGTCTTTCGAAGTGAGACACTGTTTGGATTTCAATG 592
Db 242 TGTATTAGTGTGATGTGAGTTTACAAGTCTTCGGAATGAGACGGTTCCTGGATTTCTGA 301
QY 593 TCAACTTTTATCATCAGACAGAGAACATAAATTTCAAGAACAAAGTTTCCAAAAGAACTAA 652
Db 302 CTGCTCTCTGTCAAGAAACTGGCTGTCTCTTTCACCCAGACGCTGTGAGAAGCAGCTAA 361
QY 653 TAGGTTTAGTTGTTTCTTACCAAGTATACATATAGACATACAGAGTGCATATATGACT 712
Db 362 TAGGGCTCATTGTCTCTTACAAGATACAAATACAGAACCTCTCCATTGATACATTGACT 421
QY 713 GGGACCAAGATPCCCAAGACGACCTTTTAAAGAACCGCAGGCTCTGAAGTCAAGTCTCTTAG 772
Db 422 GGTCAAGTGAAGCCCAACACACACACTTTCTTAAGCGGATGGACCCGAGATCACTATGTGG 481
QY 773 AATCTACAGAAAGCAATACAAACAGAGATCAACCGACTTGAAGCAGCGCTCTCTTGGTCA 832
Db 482 ATTACTACAGCAGCAGTATGATATTAATCTGATCGGACCTGAATCAGCCCATGCTTGTGA 541
QY 833 GCCAGCCCAAGAGAGCGGGGGCCCTGGGGGGAACACTGCCAGGGCCTGCCATGCTCATTC 892
Db 542 GTCTGTGTTAAAG---AAGAAGAGAAATGACACAGTGAAGCTCAGCTCGCCACCTGATAC 598
QY 893 CTGAGCTCTGCTATCTTTACAGGCTTAATCTGATAAATGCGTAATGATTTTAACTGATGA 952
```


Query Match		24.4%;	Score 568.6;	DB 3;	Length 3057;		
Best Local Similarity		54.9%;	Pred. No. 4.8e-142;				
Matches 1233;		Conservative	0;	Mismatches 994;	Indels 18; Gaps 5;		
QY	89	CTAACCAATTTCCGGGTGACATCCCGTCCCGAGTGGCCCTTATATCAGTATCACATTTGACT	148	Db	1520	TCCTGCCACCTGAGAACTATTTCAGAGGGGTAAACAGTTTCAGTTACATCCGAGCAATG	1579
Db	512	CAAAACGGCTTCAAGTTGAAGACTAAACAGACTGGCAACTCTACAGTACCGGGTTGACT	571	QY	1169	CAGATTGGTCCAAAGAAACAAAGAGGTGCACCATTAATTAATTAGTGTAAAGCCACTAGATACT	1228
QY	149	ATAACCCACTCATGGAAGCCAGAGACTCCGTTCCAGCTCTCTTTTCAACACGAGATC	208	Db	1580	CGGACTGGTCCAGAGACACACAGGGGTAATGCGCTCACTGATGCAAAAGATCTCTCAACACT	1639
Db	572	TTGAGCCTGAGATCTGAAACCCCTAGAGCTGTTTTGGCCCTCTGAAAGGCCACAGTGTCT	631	QY	1229	GGCTGTTGATCTATACCGGAGAAATATTAAGCAGCAGCAATTCATTGATACAAAATCTAT	1288
QY	209	TAAITGGAAAGTGTCACTGTTTGGATGGAACGATATTAATTTTAACTTAAAGACTACAGC	268	Db	1640	GGAAGATCTTCTACACAGCGCGGAGCGCCAAACAGGGCCAGGACTTTCATCAAGTCACTCG	1699
Db	632	TTCTTTGTAAGGGTTAAAGCTCGATATGGACACCATGTATTTCTTTTCAAGCTAGCAG	691	QY	1289	TTAAAGTTTACACAGCCATGGGCATGCAAAATGAGAAAAGCAATTAATGATTGAAGT---	1345
QY	269	AAAAGTTACTGAAGTTTTTAGTAGAACCCGGAATGGAGAGATGTGAGGATTAACGATCA	328	Db	1700	TACGAGTCGCAAAACCAATGGGTATGAACGTACAGGACACAGAGATAGTGAACCTCCCG	1759
Db	692	AAAAGTTAACCAACCTGAGCGCAGAGAGGAAGGATGGGTCTAAACGTGAATATCTGTGTGA	751	QY	1346	ATGACAGAACTGAAGCCTACTTTAAGAGTCTTACAGCAAAAGGTTCACAGCAGACACCCAGA	1405
QY	329	CTTTAAACAAATGAACCTCCACCTACATCACCAACTTGTTTTCAGTTCCTATAAATATTATTT	388	Db	1760	ATGATCGTACAGAGACCTTACACAGATCCCTTACAGGCACAAATCGCACAGGATACACAGA	1819
Db	752	CTCATGTGCTACTCTTAAACCCCTAGGCACCAACACCCCTGCATCTCTCAACGTCTCT	811	QY	1406	TAGTGTGTCTGTCTGTTCTCAAGTAATCGGAAGGACAAATACGATGCTATTAATAATAACC	1465
QY	389	TCAGAGGCTTTTGAAATCATGAAATTTGCAACAAATTTGACGAAATTAATTAACCCAA	448	Db	1820	TTGTGGTGGTATCCTGCCAAACGAAACAGAAAGGACCGGTACGATGCCATCAAGAGAGCT	1879
Db	812	TTAGAAGGTGTTTGAAGATGATCAATATGGAGCAAGTTGGACGCAACTACTATGATCCTA	871	QY	1466	TGTGTACAGATTGGCCCTTACCCAAAGTCAGTGTGTGTGGCCCGCAACCTTAGGCAAAACAGC	1525
QY	449	ATGACCAATTTGATATTCGAAGTACAGTTGGTGTATTTGGCTGGCTTCACTACTTCCA	508	Db	1880	GCGTGGTGACCCATCCATGCCAGCCAGGTCACTGCTCTCCAGGACCCCTGTCGAAGCAGC	1939
Db	872	CGGCCGCCATTGACATCAAGCAGCATGGGCTGCAGCTGTGGCCAGGGTTCTGTGACATCCA	931	QY	1526	AAACTGTCTAGGCCATTTGTACAAAGATTGCCCTTACAGATGAACCTCAAGATGGGAGGAG	1585
QY	509	TCCCTTCAGTATGAACCAAGCATGCTCTGCACGTAGCTTAGCCATAAAGTCTCTCGAA	568	Db	1940	AGATGCTCATGTCTGTGTAGCCCAAGATTGCCATGAGATGAACCTCAAGATGGGCGGAG	1999
Db	932	TCCCTTCAGTACAGATACAGCTGATGCTGCTCTGACATCAGCCACAGGTTCTGGTA	991	QY	1586	AGCTCTGAGGGTGGACATCCCGCTGAAGCTCGTGTATGCTGTTGGCATCGATTGTTACC	1645
QY	569	GTGAGACTGTTTGGATTTTCACTTTTATCATCAGACAGAAACATAAATTC	628	Db	2000	ACCTCTGCGAGTGCAGATCCCGCTGAGCAACCTGATGATCATCGGTATTGACAGCTACC	2059
Db	992	CCGAGCGCTCTGGAGGTCTATG---AATGATCTCTTCAACAAAGCTAGGGCGAGTTCA	1048	QY	1646	ATGACATGACAGCTGGCGGAGGTCAATCGCAGGATTTGTTGCCAGCATCAATGAAGGGA	1705
QY	629	AAGAACAGTTTCCAAAGAACTAATAGGTTTAGTGTCTTACCAAGTATTAACAATAAGA	688	Db	2060	ATGACTCTCTGACCAAGGCGCTCGCTCTGGGATTCGTGGCTTCCATGAACAAGTCTC	2119
Db	1049	AGGAGGAATCACAAAGCTCATGATTTGACACAGATTGCTCTCACAAAGTACAAACAAGA	1108	QY	1706	TGACCCGCTGGTCTCAGCTGCATATTTTCAAGATAGAGGACAGGAGCTGGTAGTGGCG	1765
QY	689	CATACAGAGTGGATGATTTGACTGGGACCAAGATCCCAAGAGCACCTTTTGAAGAAAGCCG	748	Db	2120	AAACAGCTTCTTCTCTAGTTGTGCTATTCAGCACGACAGGAGAAATTCGGTGTAAAC	2179
Db	1109	CCTACCGGTTAGATGATTCGACTTTGACACTACTCCAGGAGTACTTTTGAACCCGTT	1168	QY	1766	TCAAAGTCTCGCTGCAAGCGCTCTGAGGGCTTGGAAATAGCTGCAATGAGTACATGCCCA	1825
QY	749	ACGGCTCTGAAGTCACTTTTGAATATCTACAGGAAGCAATACAAACAGAGATCACCG	808	Db	2180	TGAGTACCTCTATGAACAATGCCCTGAGCGATCTATCAGATCAACGAGAGATTTCTGT	2239
Db	1169	CCGGACC---AGTTTCTCTATGTCGATTACTTCAAGAACTCTTATGAGAGGCTCATTCACG	1225	QY	1826	GCCGATCATCTGTGTACCGGATGGCGTAGGAGACGGCCAGCTGAAAACACTGGTGAAC	1885
QY	809	ACTTGAAGCAGCTGCTTGTGTCAGCCAGCCCAAGAGAGCGGGGCCCTGGGGGACAC	868	Db	2240	AGCGTATCATCATCTTCGTTGATGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTG	2299
Db	1226	ACGTGAACCAACCCATGCTGTGTCTCCGCCAAGAGAGGAGGAAAGAGGG-----	1280	QY	1886	ACGAGTGCACAGTCTTTTGGATTGTCTA---AAATCCATTGGTAGAGTTTACAACTTA	1942
QY	869	TGCCAGGCTGCCATGCTCATCTCTGAGCTCTGCTATCTTACAGGTCTTAATGATAAAA	928	Db	2300	ATGAGCTCAAAACAAATCAAGGACACCTTGGACAAGTCTTACCTCAAGGACGGTTTACA	2359
Db	1281	-AGTTGGTCCAGCTTACTTACCGCTGAGCTGTGCTTCTGACTGGCCCTTACGAGCAGCA	1339	QY	1943	GACTPACCGTAAATTTGGTGAAGAAAAGAGTGAACACACAGATTTTTTGTCTCAGTCTGGAG	2002
QY	929	TGCGTAATGATTTTAACTGATGAAGACTTTAGCCGTTTATPACAAAGACTAACTCCAGAGC	988	Db	2360	AGCTGGCCGTTGTGGTGTCAAGAGCGCATCAATAACCGCTTCTTTGCCAATCTGAGAG	2419
Db	1340	TGCGGCCGATTTTCAACGTCATGAAGGACATGGGCCCAACACACTCGTGTGGACCCCAAG	1399	QY	2003	GAAGACTTCAAGATPCCACTTCTCTGGAAACAGTTTATGATGTAGAGGTTTACAGACCAAGAT	2062
QY	989	AAAGCAGCGTAGTGGAGAGACTCATGATTAATTAATTAATAAAGCATATGTTTCAAA	1048	Db	2420	GAGSCCTGTCCAAACCTCCCTCCCTGGAACCGTCACTGATGATGATGATGATGATGATG	2479
Db	1400	ACAGATGTAGAACGCTCAGCGGCTTCATCAAAAAGCTCAGCAGCAATGAAGAGTGAAGA	1459	QY	2063	GGTATGACTTTTTTATCGTGAGCCAGGCTGTGAGAGTGTGTGTGTGTGTGTGTGTGTGT	2122
QY	1049	GGGAGCTTCGAGACTGGGTTTGAAGCTTTGATTTGATTTCACTTACTGTCTCTCAGGAGAA	1108	Db	2480	TCTATGACTACTTCTCATCATCAGTACAGAGTGTCCGCCAGGCTCCCGTGTCCCCCAGCT	2539
Db	1460	CTTACTAGATTCTCTGGGCGATGGAGTTTCGACAAAGAGCAGGTCAAACTGACTGGGCGTG	1519	QY	2123	ACAATGTCACTATGACAAACAGCGGCTGAAGCCAGACACACATACAGCGCTTGGACCTACA	2182
QY	1109	TTTTGCAACAGAAAAGATTACCAAGGTGGAAAAACATTTGATTGATTAATCCACAATTG	1168	Db	2540	ACAAAGTTGTCTATGACACCACTGGTCTCAAGCCAGCACCATGACAGAGTTGACCTACA	2599

Db 2600 AGCTGACCATCTCTACTTCACTGCGGGGACGGTGGTCTCTGCGCGTGCATGT 2659
Qy 2243 AGCCCAACAAGCTGGCTTTTCTTGTGGCAGAGTAATTCACAGAGCAAAATCTGTAC 2302
Db 2660 ATGCACACAAGCTGGCTTCTGTATGCGGGCAGAGCGTCCATGAAGTGCACACAGAC 2719
Qy 2303 TGTCAACCGCCTTTACTACCTTA 2327
Db 2720 TAGCCGACACCCCTCTCTATCTTA 2744

RESULT 13
AK116906
LOCUS AK116906 3458 bp mRNA linear INV 30-NOV-2002
DEFINITION Ciona intestinalis cDNA, clone:cig052a12, full insert sequence.
ACCESSION AK116906
VERSION AK116906.1 GI:23589627
KEYWORDS FLI CDNA.
SOURCE Ciona intestinalis
ORGANISM Ciona intestinalis
Eukaryota; Metazoa; Chordata; Urochordata; Ascidiacea; Enterogona;
Phlebobranchia; Cionidae; Ciona.

REFERENCE 1
AUTHORS Satou, Y., Yamada, L., Mochizuki, Y., Takatori, N., Kawashima, T.,
Sasaki, A., Hamaguchi, M., Awazu, S., Yagi, K., Sasakura, Y.,
Nakayama, A., Ishikawa, H., Inaba, K. and Satoh, N.
TITLE A cDNA resource from the basal chordate Ciona intestinalis
JOURNAL Genesis 33 (4), 153-154 (2002)
MEDLINE 22191024
PUBMED 12203911
REFERENCE 2 (bases 1 to 3458)
AUTHORS Satou, Y. and Satoh, N.
TITLE Direct Submission
JOURNAL Submitted (04-OCT-2002) Nori Satoh, Kyoto University, Department of
Zoology; Sakyo-ku, Kyoto, Kyoto 606-8502, Japan
(E-mail: satoh@ascidian.zool.kyoto-u.ac.jp, Tel:81-75-753-4095,
Fax:81-75-705-1113)
COMMENT Ciona intestinalis cDNA Project (URL:
http://ghost.zool.kyoto-u.ac.jp/indexr1.html).
FEATURES
source
1..3458
/organism="Ciona intestinalis"
/mol_type="mRNA"
/db_xref="taxon:7719"
/clone="cig052a12"

BASE COUNT 1062 a 626 c 810 g 960 t

ORIGIN
Query Match 21.6%; Score 503.2; DB 3; Length 3458;
Best Local Similarity 54.3%; Pred. No. 2.3e-124;
Matches 1225; Conservative 0; Mismatches 998; Indels 33; Gaps 9;
Qy 89 CTAAACATTTCCGGTGCATCCCGTCCCGTGGGCTTATCAGTATCATTGACT 148
Db 625 CCAACTTTTCCATTTAGTTGCACAGGACGATGGTGTCTATACAGTATAGTTGACT 684
Qy 149 ATAAACCACTGATGAGCAGCAGACTCCGTTAGCTCTTCTTTTCAACAGAGATC 208
Db 685 ATGCACCTGAGTGCACACAGAGGCGGAGAAAGGATGTTAAAGATCATGAAGCA 744
Qy 209 TAATTGGAAAGTGTCATGCTTTTGTAGTGAACGATATTTTACTTAAAGACTACAGC 268
Db 745 TCATTGGGAATGCTTTCATGTTTGTAGTGAACAAATGTTTGTACCAACCAAGAGTTGCCGG 804
Qy 269 AAAAGGTACTGAAGTTTGTAGTGAACCG---GAATGGAGAGGATGTGAGGATAAGCA 325
Db 805 AATAAGAAACAGTGGTTTACTCAAGCGTCGAACGATGAGAGCAATGTACAGCTTACGA 864
Qy 326 TCATTTAAACAATGAATTCACCTACATCAACCACTTGTGTTGAGTTCTATATTA 385
Db 865 TCGCATTAACAGAGCTTCCACCCAGTTCCCCCATCTCAATAAATAATACAACTTC 924
Qy 386 TTTTCAGGAGGCTTTTGAAAAATCATGAATTTGCAACAAATTCGACGAAATTAATTAACC 445

Db 925 TATTGAGATGGTTCCTTTGATAAGATAGCATGAAGAGTTTGAAGGAATTTTACAAATC 984
Qy 446 CAAATGACCAAAATGATATTCAAAGTACAGG-----TTGGTGAATTTGGCTGCTTCA 499
Db 985 CTTCATGTCTATTCATCTGGAATGCCAGAGGTCACACTTTGAACATATGGCCAGGATTC 1044
Qy 500 CTACTTCATCCTTTCAGTATGAAGAACAGCATCATGTCTCTGCACTGAGCTTAGCCATAAG 559
Db 1045 TCACCAGTATTCGCACTATGAGAAGAGTGTGTTTGTATGTCTGCTGAGTTTCCCATAGT 1104
Qy 560 TCCTTGAAGTGAGACTGCTTTTGGATTTCATGTTCAACTTTAT-----CATCAGACAG 613
Db 1105 TGATGCGGAAGATTCGGTGTCTGATATCTTACGCGAGATGATGGGACTGTAGGCAAC 1164
Qy 614 AAGAACAATAATTTCAAGAACAACTTTCCAAAGAACTAAATAGGTTTGTAGTTCTTACCA 673
Db 1165 GTGGAAGAGATTTTAAACAGAGATGGAATAATTTCTTGGTGGGACAAATTTCTGACAA 1224
Qy 674 AGTATAACAATAAGACATACAGAGTGGATGATATGACTGGGACAGAAATCCCAGAGCA 733
Db 1225 GATACAAACAATAAGACATACCGCATAGATGGATGGAATGGAACCTTAAATGTTAATGA 1284
Qy 734 CCTTTAGAAGCCGAGGCTCTGAAGCTCTGAACTCTTCTTAGAATCTACAGGAGCAATACA 793
Db 1285 AGTTTGACGGAATATCTGGTTCC---GTCTGATGTGGATTTATTAAGAGCAATACA 1341
Qy 794 ACCAAGAGATCACCGACTTGAAGCAGCGTCTTGGTCAGCGAGCCCAAGAGAGGCGGG 853
Db 1342 ACATAGTATCCGAGATACAAACCAACCTCTGCTTCTATCTCTCTTAAACAATCAGAGA 1401
Qy 854 GCCCTGGGGGACACTGCCAGGCGCTGCCATGCTCATTTCTTGAGCTCTGTCTATCTTACAG 913
Db 1402 TTCGAAAGGTGGATTTGGAGGTGTTTCAAT---CTTGTGCCGAGTTATGTACAGTGACAG 1458
Qy 914 GTCTAATCTGATTAATTCGTAATGATTTTAACTGATGATAAGACTTAGCCCTTCATACAA 973
Db 1459 GGCCTTACTGATGAGCTCGGGCTGACTTCAATACATGAAAAGGTGGCAGCTACACAA 1518
Qy 974 GACTAACTCCAGAGCAAGGCGGTGAAGTGGGACGACTCATTTGATTACATTTATATAAA 1033
Db 1519 AACAGGACCTTACCAAGAGAAACAAGCGCTGAAAAGCTTTATACAGCATCACCACA 1578
Qy 1034 ACATTAATGTTTCAAAGGAGCTTCGAGCTGGGTGTGAGCTTTTGAATTTCACTTACTGT 1093
Db 1579 AACTGAGTTTGAAGAGAGATTTGCAGAGTGGGGCTTAGATTTGAGGATCGACTCTTGG 1638
Qy 1094 CCTTCTCAGGAGAAATTTTGCAACAGAAAGATTCACCAAGGTGGAAAAACATTTGATT 1153
Db 1639 ATTTAAAGGGAAGAGTTCTTGAATCTGAAACCAATGTTTGGTGTATTAAGCAAGCAC 1698
Qy 1154 ACAATCCACATTTGCGAGATTTGGTCCAAAGAAACAAGAGGTGCGACCTTAATTTAGTTA 1213
Db 1699 AGAGTAGAGA---TGCACTTTGGGATCAGAGTTCCGAGACAGAGACTACTTAAAGTGA 1755
Qy 1214 AGCCACTAGATACTGGCTGTGATCTATACGCGAAGAAATTTAGACCGACCAATTCAT 1273
Db 1756 TTGATCTTTCAGGAGTGGGCCATCTCTGTTTCTCTAGAGATAAAAGATGTGCGGAGGAT 1815
Qy 1274 TGATACAAATCTATTTAAAGTTTACCGAGCTACCGCATGGGCATCAATGAGAAAAACAATA 1333
Db 1816 TTGAGAGAAATTTTAAAGTATCTGTAACTAGGATTCGGTGTGCTCGTCCGACTC 1875
Qy 1334 TGATTGAAGTGGATGACAGAACTGAAGCTTCTTAAGAGTCTTACAGCAAAAGGTC---A 1390
Db 1876 CTGTTGAGTTGGAATCGGATCAATGTTTCAGAGAGGATCCGGGATCTAATGGGAGGA 1935
Qy 1391 CAGCAGACACCCAGATAGTTCTGTCTGTGTTGTAAGTAACTCGGAAGCAAAAATCGATG 1450
Db 1936 ATCGGAAGACACAGCTGATATGTTGTATGATGCCAGCAGTAGAAGACAGATATGAAG 1995
Qy 1451 CTATTAAAAATACCTGTGTACAGATTGCCCTCCCAAGTCAGTGTGTGTCGCCCGAA 1510

Db	1396	GCATCAAGAAAGTTTGTGTGGATATGCGGTTTCCAAGTCAGGTTGTATTATCAAGCA	2055
Qy	1511	CTTTAGGCAACACAAACTGTCATGGCCATTGCTACAAAGATTGCCCTACAGATGACT	1570
Db	2056	CATTATCCAAACCAACGAGAGTAATGAGCATCGCTACAAAGATTGCAATCCAGATGAAT	2115
Qy	1571	GCAAGATGGGAGGAGAGCTCTGGAGGGTGGACATCCGCCCTCAAGCTCGTGATGATCGTTG	1630
Db	2116	GTAAGCTTGGTGGTGGAGCATGGGCTGTCAGATACCGTTAAAGAAATGTATGATCGTAG	2175
Qy	1631	GCATCGATTGTTACCATGACATGACAGCTGGCGGAGGTCAATCCAGAAATTTGTGCGCA	1690
Db	2176	GATTTCGATACATTACCAAGACTCATCACAGAAAGAAAGATCTGTTGGTGTGTTGTTGCGA	2235
Qy	1691	GCATCAATGAAGGATGACCCGCTGGTCTTCAAGCTGATATTTTTCAGATATAGAGCAGG	1750
Db	2236	GTATTAAATTCACATTACACAAAGATGGTACTCGAGTGTCACTTTTCCAAATTCAGGAGTTG	2295
Qy	1751	AGCTGGTAGATGGGCTCAAAAGTCTGCCTGCAAGCGGCTCTGAGGCTTGGAAATAGCTGCA	1810
Db	2296	AATTAAATCGATGGATTGAAACGCTGCAAGAGGGGCAATTCGAAAGTTTTCATCAGGTGA	2355
Qy	1811	ATGAGTACATCCAGCGGATATCGTGTATCCGCGATGGCGTATAGGACACGCGCAGCTGA	1870
Db	2356	ATGGGAAACTACCCGCTCGTATTTGTCATCTACAGAGATGGGTCGGTGATGGAACAGCTGG	2415
Qy	1871	AAACACTGGTGAATACGAAATGCGACAGTTTTCGATTTGATTTGCTAAA---ATCCATTGGTA	1927
Db	2416	CTGTTGTGAAGGACCATGAGATATCACAACTACTTGAATGTCTTAAGATATCCAGATATA	2475
Qy	1928	GAGGTTACAAACCCCTAGACTAAGCGGTAATGTTGGTGAAGAAAAGATGAACACCAAGATTTT	1987
Db	2476	AAGGCTTCACTCCCAATGTTCTTTCATCGTGGTTAAGAGAGATTAAACGAGTTCT	2535
Qy	1988	TGCTCACTCTGGAGGAA---GACTTCAGATTCGACTTCTGCTGGAACATTTATGATGAG	2044
Db	2536	TCGCTGCTGGTGGCGACACCAAGCAATTAATGAACGCACTCTGCTGATCATCGATGACA	2595
Qy	2045	AGGTTACAGACCCAGAAAGTATGATGATTTTATCGTGCAGCCAGCTCTGAGAGTGGTA	2104
Db	2596	CGGTACAGAACCAAGAAAGTATGATGATTTTCTTGTGTCAGCCAGTCCCTGACAGAGGCA	2655
Qy	2105	GTGTTTCTCCACACATTAACATGTCATCTATGACAAAGCGGCTGTAAGCCAGACACACA	2164
Db	2656	CAGTCAGTCCAAACGATTAACATGTTGTTATATGACACATCAACATCAACCTGATGACA	2715
Qy	2165	TACAGCGTGTGACCTTACAGCTGTCGCACTATTAACAATGCGCAGGTTGATTCGTG	2224
Db	2716	TGCAACGACTCACTTACAAAGTTGTGTACCTTTACTATAACTGGCCAGGAAACCAATTCGTG	2775
Qy	2225	TTCTGCTCCTTGGCAGTACGCCCAAGCTGGCTTTTCTGTTGGCCAGAGTATTGACA	2284
Db	2776	TACCAGCCCAATGCTATATGCGCAAGATAGCAATTTCTGTCGGCAAAAGTATTTCATG	2835
Qy	2285	GAGAGCCAAATCTGTCACTGTGTCAAACCCGCTTTACT	2320
Db	2836	CACCTGTGTCTCTACTCTAGCCGACAACTTACT	2871

RESULT 14
AY014900
LOCUS AY014900 2605 bp mRNA linear INV 02-JAN-2001
DEFINITION Strongylocentrotus purpuratus clone 6-2 seawi mRNA, partial cds.
ACCESSION AY014900
VERSION AY014900.1 GI:12007641
KEYWORDS
SOURCE Strongylocentrotus purpuratus
ORGANISM Strongylocentrotus purpuratus
Eukaryota; Metazoa; Echinodermata; Eleutherozoa; Echinozoa;
Echinoidea; Echinoidea; Echinacea; Echinacea;
Strongylocentrotidae; Strongylocentrotus.
REFERENCE 1 (bases 1 to 2605)
AUTHORS Rodriguez, A.J. and Bonder, E.M.

TITLE	Seawi - Cloning and Molecular Characterization of a Sea Urchin Homolog of Piwi
JOURNAL	Unpublished
REFERENCE	2 (bases 1 to 2605)
AUTHORS	Rodriguez, A.J. and Bonder, E.M.
TITLE	Direct Submission
JOURNAL	Submitted (28-NOV-2000) Biological Sciences, Rutgers University-Newark, 101 Warren Street Room 135, Newark, NJ 07102, USA
FEATURES	Location/Qualifiers
source	1. 2605 /organism="Strongylocentrotus purpuratus" /mol_type="mRNA" /db_xref="taxon:7668" /clone="6-2" 181..>2605 /codon_start=1 /product="seawi" /protein_id="AAG42534.1" /db_xref="GI:12007642" /translation="MDRRPGEQAPPQQAAPQEAASSGMPVPGRGRSGTTPQGGP AVFQPRGELTSQAKIALGASGGRVGAQDRRRRDEIELEPIITRENFKQAI AGDKIALIANGFKLTKPDWLYQYRDFEILNPRFALLKGLSALGKGLTLM DTMYSFLKAEKVTNLSEKSGSNVNICVHTATLNPAPNTLHLYNLFRRCLKM NMEQVGRNYDPTAADI KOHGLQWPGFVTSILOEYDMLLSIDISHKVLRTQVWE VMDLENKARGREKERITKLMIGIOLVLTNNKTYRVDIDEDTPTATPETSQPV YVDYFKSIRVTHDYNQMLVSRPKRBEKGVGPAYLPPELFLTGLSDDWRADEN VMDMGHTRVGFQDRCTLSGIFKLSSNEBKVYLDLSWMEFDEKQVLTGRLVPP EKLFQKQKQSYNPSNADSRDRTGNALTDKILNNKI FYRDRANRGQDFIKSLVR VAMPNMGVRGPEIPELDDRTETYSLSQAQADTQIVVILPTTKRDYDAIKKT CVYTHPCPSOIVSRITLSKQMLMSVATKIAMQNCNMGGDLWRVETPLNLMITIGID SYHDSLTGKRSVLGFSVAMNKSOTSPSCAFOHAQCFGANLSLTMNALKRYOIN EKPERILIFRDGVDGSOVNLVVDYELKOLKDLKIVYPOGTVHKHAAVAVVKEINR FFANRLGGLSNPPPGIVDDVTKPHLYDFILISQSVRSQSVSFTSNVYDTGLKE DHMDVITYGI"
CDS	BASE COUNT 718 a 695 c 686 g 506 t Query Match 21.5%; Score 499.4; DB 3; Length 2605; Best Local Similarity 54.1%; Pred. No. 2.4e-123; Matches 1135; Conservative 0; Mismatches 946; Indels 18; Gaps 5; Qy 89 CTAACCAATTTCCGGCTGACATCCCGTCCCGAGTGGGCTTATATACAGTATCACATTTGACT 148 Db 512 CAACAGCGCTTCAAGTTGAGACATAAACAGACTGGCACTACACAGTACCGGTTGACT 571 Qy 149 ATAAACCACTGATGGAAGCCAGAGACTCCGTTGAGCTCTTCTTTTCAACAGCAAGATC 208 Db 572 TTGAGCCTGAGATCCTGAAACCTAGAGCTGTTTGGCCCTCCTGAAAGGCCACAGTGTCTC 631 Qy 209 TAATTGGAAAGTGCATGCTTTTTCATGGAACGATATATTTTACCTAAAAGACTACAGC 268 Db 632 TTCTTGTGTAAGGTTTAAAGCTTCGATATGACACCAATGATTTCTCTTTTCAAGCTAGCAG 691 Qy 269 AAAAGGTTTACTGAAAGTTTTCATGTAAGCCCGGAATGGAGAGATGTCGAGGATAACGATCA 328 Db 692 AAAAGGTTAAACCACTGAGCCGACAGAGAGGATGGTCTAAAGTGAATATCTGTGTA 751 Qy 329 CTTTAAACAATGAACTTCCACCTCATCACCAACTTGTGTGAGTTCTATTAATATTTT 388 Db 752 CTGATGTGCTACTCTTAAACCCCATGGCACCAACACCCCTGATCTCTACAAACGCTCTCT 811 Qy 389 TCAGAGGCTTTTGAATATCATGAAATTCGAACAAATGGACGAATATATATATACCCAA 448 Db 812 TTAGAAGGTGTTTGAAGATGATCAATATGGAGCAAGTTGGACGCAACTACTATGATCCTA 871 Qy 449 ATGACCCAAATGATATTCCTCAAGTCAACAGTTGGTGGATTTGGCCCTGGCTTCACTACTTCCA 508 Db 872 CGGCCGCCATGACATCAAGCAGCATGGCTGACGTGTGGCCAGGGTTCGTGATATCCA 931 Qy 509 TCCTTCAGTATGAAAACAGACATCATGCTCTGCACTGACGTTAGCCATAAAGTCCTTCGAA 568 Db 932 TCCTTCAGTACGAGTACGAGCTGATGCTGCTGACATCAGCCACAGGTTCTGCGTA 991

Qy 569 GTGAGACTGTTTGTGATTTCAATGTTCACTTTTATCATCAGACAGAAACATAAAATTC 628
Db |||||
Qy 992 CCCAGCGTCTGGGAGGTCTATG---AATGATCTCTTCAACAAAGCTAGGGCAGGTTCA 1048
Db |||||
Qy 629 AAGACACAGTTTCCAAAGNACTAATAGGTTTGTCTTACCAAGTATACATNAGA 688
Db |||||
Qy 1049 AGGAGGAATCACAAGCTCATGATGACAGATTGTCTCACAAGATACACACAGA 1108
Db |||||
Qy 689 CATACAGAGTGGATGATATTGACTGGGACCCAGAAATCCCAAGAGCACCTTTAAGAAAGCCG 748
Db |||||
Qy 1109 CCTACCGGTAGATGATATCGACTTTGACACTTACTCCAGCGGATATTTTGGACCCGTT 1168
Qy 749 ACGGCTCTGAAGTACGCTTTTGAATATCTACAGGAAGCAATACCAAGAGATCACCG 808
Db |||||
Qy 1169 CCGGACC---AGTTTCTTATGTCGATTTACTTCAAGAAGTCTTATGAGAGGTTCAATTCAG 1225
Qy 809 ACTTGAAGCAGCTCTGTTTGTGTCAGCCAGCCCAAGAGAGCGGGCCCTGGGGGACAC 868
Db |||||
Qy 1226 ACGTGAACCAACCAATGCTGTGTCTGCCCAAGAGAGGAGGAAAGAGGG----- 1280
Qy 869 TCCAGGSCCTGCCATGCTCATCTCTGAGCTCTGCTATCTTACAGGTCTAACTGATAAAA 928
Db |||||
Qy 1281 -AGTTGTCTCAGCCTACCTACCGCTGAGCTGTCTTCTGACTTGGCCTTAGCGAGACA 1339
Qy 929 TCGGTATGATTTTAACTGATGAAGAGCTTAGCCGTTTCATCAAGACTAACTCCAGAGC 988
Db |||||
Qy 1340 TCGCGCGCGATTTCAACGTCATGAGGACATGGGCCAACACACTCGTGTGGACCCCAAG 1399
Qy 989 AAAGSCAGCTGAAGTGGAGGACCTCATTTGATTTACATTCATAAAGCAATATGTTCAA 1048
Db |||||
Qy 1400 ACAGATGTAGACGCTCAGCGCTTCATCAAAAGCTCAGCAGCAATGAAGAGTGAAGA 1459
Qy 1049 GGGAGCTTCGAGAGTGGGTTTGAGCTTTGATTTCAACTTACTGTCTTCAGAGAGAA 1108
Db |||||
Qy 1460 CTTACTAGATTCCTGGGCGATGAGTTCGACAGAGCAGGTCAAACCTGACTGGCGTG 1519
Qy 1109 TTTTCAACAGAAAGATTCCAGAGGTGGAAACATTTGATTACATCCACAATTTG 1168
Db |||||
Qy 1520 TCCTGCCACCTCAGAAACTATTTCCAGAGGGGTAAACAGTTCAGTTTACAAATCCGAGCAATG 1579
Qy 1169 CAGATTGTCACAAACACAGAGGTGACCAATTAATTAGTTTAAAGCCACTAGTAACT 1228
Db |||||
Qy 1580 CGGACTGTTCAGAGACACCGAGGTATGCGCTCACTGATGCAAGATCTCCAACT 1639
Qy 1229 GGTCTTGTATCTATACGGCAAGAAATTATGAGCAGCCAAATTCATTTGATACAAATCTAT 1288
Db |||||
Qy 1640 GGAAGATCTTCTACACACCGCGGACGCCAACAGGGGCCAGGACTTCATCAAGTCACTCG 1699
Qy 1289 TTAAGTTTACACAGCCATGGGCATGCAAAATGAGAAAGCAATATGATTGAAGTG---G 1345
Db |||||
Qy 1700 TACGATCGCAAAACCCAAATGGGTATGAACGTACGAGGACCCAGAGATAGTGGAACTTCCCG 1759
Qy 1346 ATGACAGAACTGAAGCCTTACTTAAAGTCTTACAGCAAAAGTCCACAGCAGACCCAGA 1405
Db |||||
Qy 1760 ATGATCGTACAGACCTTACAGATCCCTTACAGGCACAAATCCGACAGATACACAGA 1819
Qy 1406 TAGTTGTCTGTCTGTGTCAAGTAAATGGAAGGACAAATACGATGCTATTAATAATACC 1465
Db |||||
Qy 1820 TTGTGGTGTGATCTCTGCAACGAAACAGAAAGGACCGGTACGATGCCATCAAGAGAGCT 1879
Qy 1466 TGTGTACAGATTGCCCTACCCCAAGTCTAGTGTGTGGTGGCCGCAACCTTAGGCAACAGC 1525
Db |||||
Qy 1880 GCGTGGTGACCATCCATGCCCAAGCCAGGTCTCTCCAGGACCCCTTCCAGCAGC 1939
Qy 1526 AAATGTCTATGGCCATGCTCAAGAGATTGCCCTACAGATGAATGCAAGATGGAGGAG 1585
Db |||||
Qy 1940 AGATGCTCATGTCTGTAGCCACCAAGATTGCCATGCAGATGAATGCAAGATGGCGGAG 1999
Qy 1586 AGCTCTGGAGGTGGACATCCCGCTTGAAGCTCGTGTGATGATCGTTGGCATCGATTGTTACC 1645
Db |||||
Qy 2000 ACCITGCGGAGTCGAGATCCCGCTGAGCAACCTGATCATCGTATTGACAGCTACC 2059

Qy 1646 ATGACATGACAGCTGGCGGAGGTCAATTCGAGGATTTGTTGCCAGCATCAATGAAGGA 1705
Db |||||
Qy 2060 ATGACTCTCTGACCAAGGCGGCTCCGTCCTGGGATTCGTGGCTTCCATGAACAGTCTC 2119
Qy 1706 TGACCCCTGTTTCTCAGCGTGCATATTTTCAAGATAGAGCAGAGCTGTGTAGTGGG 1765
Db |||||
Qy 2120 AAACAGCTTCTTCTCTAGTTGTGCAATTCAGCAGCAGCAGGAGAAATTCGTGCTAAC 2179
Qy 1766 TCAAAGTCTGCTTCAAGCGGCTCTGAGGGCTTGAATAGTCAATGAGTACATGCCCCA 1825
Db |||||
Qy 2180 TGAGTACCTCATGAACAATGCGCTGAGCGATATCATGATCAACGAGAGTTTCTG 2239
Qy 1826 GCCGATCATCTGTGTAACCGGATGCGGTAGAGACGGCCAGCTGAAAACACTGTGTAACT 1885
Db |||||
Qy 2240 AGCTATCATCATCTTCCGTGTGTGTGTGTCAGACCAAGTCAATCTGTCTTGA 2299
Qy 1886 ACGAAGTGCCACAGTTTGTGA---TTGTCTAAATCCATTGGTAGAGTTTACACCTA 1942
Db |||||
Qy 2300 ATGAGCTCAAAACAAATCAAGGACACCTTGGACAAAGTCTTACCTCAAGGGACGGTTACA 2359
Qy 1943 GACTAAGCGTAATTTGTGTGAAGAAAGAGTGAACACAGATTTTGTCTCAGTCTGAG 2002
Db |||||
Qy 2360 AGCTGGCGTGTGTGTCTCAAGAGCGCATCAATAACCGCTTCTTGGCAATCTGAGAG 2419
Qy 2003 GAAGACTTCAAGATCCATCTCTCGGAACAGTTATTGATGTAGAGTTACCGAGACAGAA 2062
Db |||||
Qy 2420 GAGGCTGTCCAAACCTTCCCTCGGAACCGTCTCATGATGATGTCGTCAACCAAGCCACC 2479
Qy 2063 GGTATGACTTTTATCTGTCGAGCGCGCTGTGAGAGTGTGTGTTTCTCCACACATT 2122
Db |||||
Qy 2480 TCTATGACTACTTCTATCATCAGTCAGAGTTCGCCAGGGCTCCGTGTCCCGCCAGTCT 2539
Qy 2123 ACATGTCTATATGACAAACAGCGGCTGAAAGCAGACACATACAGCGGTGACCTAC 2181
Db |||||
Qy 2540 ACHAGTTGTCTATGACACCACTGTCTCAAGCAGACCACTGACAGAGTTGACCTAC 2598

RESULT 15
AB032605

LOCUS AB032605 4913 bp mRNA linear ROD 04-OCT-2001
DEFINITION Mus musculus mRNA for MILI, complete cds.
ACCESSION AB032605
VERSION AB032605.1 GI:7416112
KEYWORDS MILI (Mili like).
SOURCE Mus musculus (house mouse)
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE 1
AUTHORS Kuramochi-Miyagawa,S., Kimura,T., Yomogida,K., Kuroiwa,A.,
Tadokoro,Y., Fujita,Y., Sato,M., Matsuda,Y. and Nakano,T.
TITLE Two mouse plwi-related genes: miwi and mili
JOURNAL Mech. Dev. 108 (1-2), 121-133 (2001)
MEDLINE 21463379
PubMed 11578866
REFERENCE 2 (bases 1 to 4913)
AUTHORS Miyagawa,S.K. and Nakano,T.
TITLE Direct Submission
JOURNAL Submitted (21-SEP-1999) Satomi K Miyagawa, Research Institute for
Microbial Diseases, Department of Molecular Cell Biology; Yamadaoka
3-1, Suita, Osaka 565-0871, Japan
(E-mail:smiyagaw@biken.osaka-u.ac.jp, Tel:81-6-6879-8362,
Fax:81-6-6879-8362)
COMMENT Sequence updated (04-Dec-1999).
FEATURES
Location/Qualifiers
1..4913
/organism="Mus musculus"
/mol_type="mRNA"
/db_xref="taxon:10090"
/map="14D3"
/tissue_lib="testis"
1..4913
/gene="mili"

gene

CDS

198...3113
/gene="mili"
/note="Mili like protein
piwi family"
/codon_start=1
/product="Mili"
/protein_id="BAA93706.1"
/db_xref="GI:7416113"
/translation="NDVVRPLFRGTPVHPQCVMPGCCWQAPRPLEPAWGRAGPAG
GLVFRKPESSLPQVQKDSGLVNSFRMGGLDTAFPPSPKREVPLGLRGLVGRGL
SANVVRKDERPRLSPDPVIAAGDSKLAESVGRMLGRSSVSLPLGRRAAS
IGRWKDPSPALPQPPALSPPTSLSDAPPVPLVLMERKEKLVQGSK
GTQSLGLNLKIACHNEAVQYHVTFSPSVECKSMRGLKHDSQVNTVAFDGS
LVLPKLQOVVELSKQRTDAEISIKIQLKILEPCSDLICI PYNVVRERWMLDM
KLVRNFDPTPSAMVLOKRLIQNPGVYASIRRDGGLFLLADVSHKVRINDSVLDV
HAIYQONKREFOBCKSLVLSI VITRYNNRTYRIDVDWNTKPKDSFVMSDGBEIT
LEYSKNTYITKEDDDQLHFRSPQRNNHMLLKEILLPELSPFTGIPERKGL
FRAMKDTQQNLSPKQHGHALECLLQRIQSONETASNELRWGLSLKHVDHKEIRLL
PMERINLRNTSFVSDNLNWKVEYTRDASI IUTPMHFWALFYPKRAMDOARELVNMLE
KIAGPIGRIISPPAWVELKDDRIETIYRTIOSLLGVGKIQMVVCIIMGTFDDLYGAI
KMLCVQSPVSOVINRTIQPTRLRSVAQKILLQNMCKLGLGELGWIDIPKLQMLVI
GNDVHDSRGRSVGVGASINLTITKWSRVVFMPEHBI VDSLKLCILGVSLKKYI
EVNCLPEKIVVYDSDGGLQTVANYEIPQLQKCFEAFDNYHPKMWVFWVKKIST
NLYLAAPHFTVPSPTVVDHTITSCWVDYFLLAHVRQCGGPTPIYICVLTANILS
PDHMORLTFKLCNWNWPGTIRVPAPCKYAKHLAFLSLGQILHHEPAIQCGNLFEL"
BASE COUNT 1260 a 1157 c 1260 g 1236 t
ORIGIN

Query Match 19.8%; Score 460.6; DB 10; Length 4913;
Best Local Similarity 52.4%; Pred. No. 7.8e-113;
Matches 1162; Conservative 0; Mismatches 1034; Indels 21; Gaps 6;
Qy 124 GCCTATATCAGTATCATTGACTATTAACCCAGTGGAGGACGAGACTCGGTCA 183
Db 906 GCAGTTTATCAGTACCATGTGACTTTCAGCCCCAGTGTGGAATGCAAAAGATGAGGTT 965
Qy 184 GCTCTCTTTTTCACACGAAAGATCTAATTTGGAAGTGTCTATGCTTTTGTGGAGACGATA 243
Db 966 GGCAATTTGAGGACCACCATCTGTCTACTGGAAAGCTCACTGCTTTTGTGGCTCTATT 1025
Qy 244 TTAATTTTACCTAAAGACTACGAAAGGTTACTGAAGTTTTPAGAACCCGGAAT 303
Db 1026 CTTTATCTCTCTGTTAAGCTTCAACAAGTGTGTAGTTTAAAAAGTCAGAGAAACTGAC 1085
Qy 304 GGAGAGATGTGAGATACGATCACTTTACAAATGAATCTCCACTACATCACCRACT 363
Db 1086 GATCGGAGATCAGTATCAAGATCAGCTGACAAAGATCCCTGGAGCCGTGTCTGACCTG 1145
Qy 364 TGTGTGCAAGTCTTATAATATTATTTTCAGGAGCGCTTTTGAATAATCATGAATTTGCAACA 423
Db 1146 TGCATCCCTTCTACAAATGTCTCTCCGCGGGTAATGAAACTTCTGGATATGAGCTT 1205
Qy 424 ATTGGAAGAAATATTATAACCAATGACCAATGTATATTCAGATTCACAGGTTGGTG 483
Db 1206 GTGGGGAAGAACTTATGACCTCAAGTGTGCAAGTGTGCTGAGCAACACAGAGTTGAG 1265
Qy 484 ATTTGGCTGCTTCACTACTTCCATCTCTCAGTATGATGAAACAGCATCATCTCTGCACT 543
Db 1266 ATCTGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1325
Qy 544 GACGTTAGCCATAAAGTCTCTGAAAGTGAAGTGTGTTTGGATTTTCAATGTTCAACTTTAT 603
Db 1326 GATGCTCTCATAAGGTCATTGGAACGACTCTGTGCTGGATGTGATGATGATGATGATGAT 1385
Qy 604 CATCAGACAGAAACATAAATTTCAAGAACAGTTTCCAAAGAACTAATAGTTTACTT 663
Db 1386 CAGCAGAACAAAGGAC---ACTTCAGGACGAGTGCAGCAAGCTTCTGCTGGCAGCAATT 1442
Qy 664 GTTCTTACCAAGTATAAATAAGACATACAGAGTGGATGATATTGACTGGACCAAGAA 723
Db 1443 GTCACTACGCGCTACAAATCTGTACCTACCGAATCGATGATGATGATGATGATGATGAT 1502
Qy 724 CCCAAGAGCACCTTTTAAGAAAGCCGAGCGCTCTGAAGTCAAGTCTTCTTAGAATACTACAGG 783

Db 1503 CCTAAGACAGCTTTGTCTGTCTGGACGGGAAGAAATCACATTCCTGGAATCTACAGC 1562
Qy 784 AAGCAATACAAACAGAGATCACCGACTTGAAGCAGCCTGTCTTTGGTCAGCACGCCCA-- 841
Db 1563 AAAAACTATGGGATCACAGTCAAGGAAGATGACAGCGCGTGTCTGATCCACCGGCCAAGT 1622
Qy 842 -AGNAGAGCGGGCCCTCGGGGGACACTGCCAGGCGCTGCCATGCTCATTCCTCAGGCTC 900
Db 1623 GAGAGACAGAATACCAATGGCATGTCTGCTGAAGGGCGAGATCTCTGCTGCTCCCGAGCTC 1682
Qy 901 TGCCTATCTTACAGCTCTAACTGTATAAATCGTAAATGATTTTAACTGTGATGAAAGACTTA 960
Db 1683 TCCTTCATGACGGGGATCCCTGCAAGATGAAGAGAGCTTCAGGGCCATGAAGAGCTTG 1742
Qy 961 GCGCTTCATCAAGACTAACTCCAGACAAAGGAGCGGTGAAGTGGAGCGACTCATGAT 1020
Db 1743 ACTCAGCAGATTAACTTGAGCCCCCAGCAGCACCCAGCGTGTCTTGAATGCTCTGCTGAG 1802
Qy 1021 TACATTCATAAAGACGATAATGTTCAAAGGAGCTTCGAGACTGGGGTTCAGCTTTGAT 1080
Db 1803 AGAATTTTCAAAAACGAGACAGCCAGCAATGAGCTGACCCGCTGGGGCTCAGTCTGCAT 1862
Qy 1081 TCCAACTTACTGTCTCTCAGGAAGAAATTTTCAAAACAGAAAAGATTCAACAAAGTGGGA 1140
Db 1863 AAGATGCTCCACAAGATTGAGGTGCGCTTCTGCCATGAGAGGATCACTTAAGGAAC 1922
Qy 1141 AAAACATTTGATTAATCCAAATTTGAGATTTGGTCTTCCAAAAGAAACAGAGGTGACCA 1200
Db 1923 ACTTCATTTG---TCATATCGGAGGACTGAACTGAGTAAAGGAAGTACACAGAGATGCT 1979
Qy 1201 TTAATAGTGTTAAGCCACTAGATAACTGGCTGTGTATCTATACGCGAAGAAATTTAGAA 1260
Db 1980 TCCATTTCTAATATTCCTGCTGCTTCTGGGCACTCTTTTATCCAAAGAGAGCAATGGAC 2039
Qy 1261 GCAGCCAAATTCATTGATACAAAATCTATTTAAAGTTACACAGCAGCATGGGCATGCAATG 1320
Db 2040 CAAGCCAGAGAACTGGTTAACTGTTGAAAAGATTGCGGGGCCCAITGGCATGCGCAT 2099
Qy 1321 AGAAAAGCAATATGATTTGAAGT---GGATGACAGAACTGAAGCTTCTTAAGAGTCTTA 1377
Db 2100 AGCCCCCAGCTGGTGTGAGCTGAAGGATGACCGAATAGAGACTATATCAGGACCAAT 2159
Qy 1378 CAG-----CAAAAGGTCACAGACAGACCCAGATAGTGTCTGCTCTGCTCAAGTAAT 1431
Db 2160 CAGTCTTCTTGGAGTTGAGGGAAGATCAAAATGCTGCTTTCATCATCATGCGGACCA 2219
Qy 1432 CGGAAGGACAAATACGATGCTATTATAAAAATACCTGTGTACAGATTGCTTACCCCAAGT 1491
Db 2220 CGTATGATCTCTATGAGCCATCAAGAGCTGTGCTGCTGCTGCTGCTGCTGCTGCTGCT 2279
Qy 1492 CAGTGTGTGGTGGCCGAACTTAGGCAAAACAGCAAACTGTCTATGGCCATTTGCTACAAAG 1551
Db 2280 CAGGTCTATCAATGTGCGCAACCAATTTGTCAGCCCAAGGCTTCGGAGCGTGGCTCAGAAA 2339
Qy 1552 ATTGCGCTACAGATGAATGCAAGATGAGGAGGAGCTCTGGAGGGTGGACATCCCGCTG 1611
Db 2340 ATTTTACTTCAATGAACTGTAAACTGGTGTGAGCTCTGGGGAGTGGATATTTCGCTG 2399
Qy 1612 AAGTCTGTGATGATGCTTGGCATCGATTGTTTACCATGACATGACAGCTGGGCGGAGTCA 1671
Db 2400 AAACAACTAATGTTGATTTGGAATGGATGTGTACATGATGCCAGCAGAGGATCGGCTCT 2459
Qy 1672 ATCGCAGATTTTGTTCAGCATCAATGAAGGGATGACCCGCTGGTGTCTCAGCTGATTA 1731
Db 2460 GTGTGCGGCTTCGTGGCCAGCATAAATCTCACACTCACCAAAATGTTACTCGAGGGTGGTG 2519
Qy 1732 TTTTCAGATGAGACAGAGAGCTGGTATGAGGCTCAAGTCTGCTGCTGCAAGCGGCTCTG 1791
Db 2520 TTCCAGATGCCCATCATCAGGAGATTGTGACAGCGCTGAAGCTCTGCTGCTGGTGGTTCCTTG 2579
Qy 1792 AGGCTTTGGAAATAGTGCATATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1851

Db 2580 AAAAAGTATATAGAGTGAACCATTTGCTCTCCAGAGAAAATTTGGTGTACCGAGATGGA 2639
Qy 1852 GTAGAGAGCGCCAGCTGAAAACTAGTGGAATACGAAGTCCCAAGTTTTTGGATTGT 1911
Db 2640 GTGCTGATGGCCAGCTAAAGACAGTTGCCAACTACGAGATCCCTCAGCTGCAGAAGTGT 2699
Qy 1912 CTAAATCCATTGGTAGAGTTTACACCTAGACTAAACGGTAATTTGGTGAAGAAAAGA 1971
Db 2700 TTGAAGGCTTTGATAAC--TACCACCCCAAGATGGTGGTGTGTAGTTCCAGAAGAAA 2756
Qy 1972 GTGAACACAGATTTTTTGTCTCAGTCTGGAGGAGACTTCAGAATCCACTTCTCTGGAACA 2031
Db 2757 ATCAGCACCAATCTGTACCTTGTCTCTGATCACTTCGTAACCCCTCCCCCGGACT 2816
Qy 2032 GTTATTGATGTAGAGTTACCAAGACAGAAATGGTATGACTTTTTTATCGTAGAGCAGGCT 2091
Db 2817 GTGGTTGATCATACCATAACCAAGCTGTGAGTGGTGGATTTCTACCTTCTTGCCCATCAT 2876
Qy 2092 GTGAGAAGTGTAGTGTCTTCTCCACACATTAATGATCATCTATGACAACAGCGGCGCTG 2151
Db 2877 GTGCCAGAGGCTGTGGCATACCTACACACTACATCTGTGTCTGAACACTGCAAAATCTG 2936
Qy 2152 AAGCCAGACCAATACAGCGCTTGACCTACAAAGCTGTGCCACATCTATTACAACTGGCCA 2211
Db 2937 AGCCCTGATCATATGCAGAGTTGACTTCAACTATGCCACATGTACTGGATTTGGCT 2996
Qy 2212 GGTGTCTATCGTGTCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 2271
Db 2997 GGTACCATCCGAGTTCCAGCTCTTTCGAAGTATGCCCAAGCTAGCTTTCTCTGTCGGGA 3056
Qy 2272 CAGAGTATTCAGAGAGCGCAATCTGTACATGTCAACCGGCTTTACTACCTCTAA 2328
Db 3057 CAGATTTTGCATCATGAGCCAGCCATCCAGCTGTGTGGGAACCTGTCTCTCTGTAA 3113

Search completed: December 6, 2003, 14:28:52
Job time : 8522 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: December 6, 2003, 08:58:52 ; Search time 617 Seconds ;
(without alignments)

10185.238 Million cell updates/sec

Title: US-10-043-774B-1

Perfect score: 2328

Sequence: 1 atgattcttggtgaacac.....accgcctttactacctctaa 2328

Scoring table: IDENTITY_NUC

Gapop 10.0 , Gapext 1.0

Searched: 2552756 seqs, 1349719017 residues

Total number of hits satisfying chosen parameters: 5105512

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

N Geneseq 19Jun03:*

- 1: /SIDSL1/gcgdata/geneseq/geneseqn-emb1/NA1980.DAT.*
- 2: /SIDSL1/gcgdata/geneseq/geneseqn-emb1/NA1981.DAT.*
- 3: /SIDSL1/gcgdata/geneseq/geneseqn-emb1/NA1982.DAT.*
- 4: /SIDSL1/gcgdata/geneseq/geneseqn-emb1/NA1983.DAT.*
- 5: /SIDSL1/gcgdata/geneseq/geneseqn-emb1/NA1984.DAT.*
- 6: /SIDSL1/gcgdata/geneseq/geneseqn-emb1/NA1985.DAT.*
- 7: /SIDSL1/gcgdata/geneseq/geneseqn-emb1/NA1986.DAT.*
- 8: /SIDSL1/gcgdata/geneseq/geneseqn-emb1/NA1987.DAT.*
- 9: /SIDSL1/gcgdata/geneseq/geneseqn-emb1/NA1988.DAT.*
- 10: /SIDSL1/gcgdata/geneseq/geneseqn-emb1/NA1989.DAT.*
- 11: /SIDSL1/gcgdata/geneseq/geneseqn-emb1/NA1990.DAT.*
- 12: /SIDSL1/gcgdata/geneseq/geneseqn-emb1/NA1991.DAT.*
- 13: /SIDSL1/gcgdata/geneseq/geneseqn-emb1/NA1992.DAT.*
- 14: /SIDSL1/gcgdata/geneseq/geneseqn-emb1/NA1993.DAT.*
- 15: /SIDSL1/gcgdata/geneseq/geneseqn-emb1/NA1994.DAT.*
- 16: /SIDSL1/gcgdata/geneseq/geneseqn-emb1/NA1995.DAT.*
- 17: /SIDSL1/gcgdata/geneseq/geneseqn-emb1/NA1996.DAT.*
- 18: /SIDSL1/gcgdata/geneseq/geneseqn-emb1/NA1997.DAT.*
- 19: /SIDSL1/gcgdata/geneseq/geneseqn-emb1/NA1998.DAT.*
- 20: /SIDSL1/gcgdata/geneseq/geneseqn-emb1/NA1999.DAT.*
- 21: /SIDSL1/gcgdata/geneseq/geneseqn-emb1/NA2000.DAT.*
- 22: /SIDSL1/gcgdata/geneseq/geneseqn-emb1/NA2001A.DAT.*
- 23: /SIDSL1/gcgdata/geneseq/geneseqn-emb1/NA2001B.DAT.*
- 24: /SIDSL1/gcgdata/geneseq/geneseqn-emb1/NA2002.DAT.*
- 25: /SIDSL1/gcgdata/geneseq/geneseqn-emb1/NA2003.DAT.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	ID	Description
1	2311	99.3	3472	21	AAA07588 Human piwi gene, d
2	1680.6	72.2	4064	21	AAA07587 Mouse piwi gene, d
3	1569	67.4	2292	24	ABV89287 Human colon cancer
4	458.4	19.7	2009	24	ABQ61109 Human mRNA sequenc
5	395.4	17.0	423	21	AAF21880 Human breast and o
6	381	16.4	3649	22	AAAS13630 cDNA sequence enco
7	365.4	15.7	367	25	ABZ19511 Group III cDNA can
8	364	15.6	367	25	ABZ20005 Group III cDNA can

9	330	14.2	2272	22	AAH15360 Human cDNA sequenc
10	310.6	13.3	2838	23	ABL10571 Drosophila melanog
11	286.4	12.3	300	20	ABL13404 Human gene express
12	263.4	11.3	3524	23	ABL10535 Drosophila melanog
13	256.6	11.0	3047	21	AAA07586 Drosophila piwi ge
14	237.2	10.2	1733	22	AAH14032 Human cDNA sequenc
15	207.4	8.9	209	25	ABZ19102 Group III cDNA can
16	184	7.9	5671	23	ABL10570 Drosophila melanog
17	171.8	7.4	548	24	ABN62610 Human cancer relat
18	149.2	6.4	1026	23	ABN70144 DNA encoding novel
19	138	5.9	7667	23	ABL10534 Drosophila melanog
20	113.2	4.9	786	22	AAH05561 Human cDNA clone (
21	111.2	4.8	2914	22	AAK53425 Human polynucleoti
22	111.2	4.8	2914	25	ABX13625 Human cytokine or
23	111.2	4.8	3050	22	AAH16058 Human cDNA sequenc
24	111.2	4.8	3736	25	ABX34752 Human mddt cDNA SE
25	104.2	4.5	325	22	AAH4876 Human polynucleoti
26	95.6	4.1	444	23	AAH57327 cDNA #3 encoding p
27	91.2	3.9	601	23	AAH70143 DNA encoding novel
28	90	3.9	646	23	AAH77101 Human cDNA sequenc
29	86.6	3.7	3996	22	AAH14510 Translation initia
30	83.2	3.6	2951	22	AAH24562 Human cDNA clone (
31	82	3.5	642	22	AAH07789 Human cDNA sequenc
32	82	3.5	1968	22	AAH14966 Human coding sequ
33	81.8	3.5	3011	24	ABQ99304 Human cDNA encodin
34	81.8	3.5	3580	24	ABS78717 Human cDNA clone (
35	79.8	3.4	516	22	AAH05278 Human gene express
36	79	3.4	734	20	AAZ15059 Human immune/haema
37	78.8	3.4	771	22	AAK58246 Human genome-deriv
38	78.6	3.4	213	24	ABS17735 Arabidopsis thalia
39	69.6	3.0	1158	21	AAK50998 Arabidopsis thalia
40	69.6	3.0	3195	21	AAK50992 Arabidopsis thalia
41	69.6	3.0	3336	21	AAK47079 Arabidopsis thalia
42	68	2.9	364	22	ABA56879 Human foetal liver
43	68	2.9	384	22	ABA26498 Probe #4964 for ge
44	68	2.9	364	24	ABS05149 Human genome-deriv
45	64.2	2.8	166	22	ABA69444 Human foetal liver

ALIGNMENTS

RESULT 1

AAA07588 ID AAA07588 standard; DNA; 3472 BP.

XX AC AAA07588;

XX 29-AUG-2000 (first entry)

XX Human piwi gene, designated hiwi.

XX Piwi family protein; piwi; miwi; hiwi; gene therapy; tissue dystrophy;
KW anaemia; immunodeficiency; male infertility; human; ds.

XX Homo sapiens.

XX Key Location/Qualifiers

FT CDS 167..2752

FT /*tag= a

FT /product= hiwi

FT /transl_except= (pos:392..394; aa:Xaa)

FT /transl_except= (pos:1073..1075; aa:Xaa)

FT /transl_except= (pos:2369..2370; aa:Xaa)

FT /note= "Xaa= leu or Ile"

XX WO200032039-A1.

XX 08-JUN-2000.

XX 03-DEC-1999; 99WO-US28764.

XX 04-DEC-1998; 98US-0110901.

XX (UYDU-) UNIV DUKE.
 PA Lin H;
 XX WPI: 2000-412085/35.
 DR P-PSDB; AA90235.
 XX Piwi family nucleic acids, polypeptides, and antibodies, useful in gene
 PT therapy of diseases such as cancer and in various research and
 PT diagnostic applications -
 XX Claim 19; Page 189-194; 201pp; English.
 XX This sequence encodes the human piwi family protein, designated
 CC hiwi. The piwi family nucleic acids and polypeptides are used in gene
 CC therapy of diseases such as cancer and also in various research and
 CC diagnostic applications. The sequences can also be used to treat
 CC tissue dystrophy, anaemia, immunodeficiency, and male infertility.
 XX Sequence 3472 BP; 1044 A; 694 C; 796 G; 935 T; 3 other;
 SQ Query Match 99.3%; Score 2311; DB 21; Length 3472;
 Best Local Similarity 99.7%; Pred. No. 0;
 Matches 2314; Conservative 0; Mismatches 7; Indels 0; Gaps 0;

QY 8 TTGGTGTGAACACAGGCGAGACTTAGACCATGTTAAAGATCAAAAACAGGTTCTTCAG 67
 DB 432 TTGGTGTGAATCAAGCGAGAACCTAGACCATGTTAAAGATCAAAAACAGGTTCTTCAG 491

QY 68 GCATTATAGTAGGTTAAGCACTAACCATTTCCGGCTGACATCCCGTCCCGAGTGGGCT 127
 DB 432 GCATTATAGTAGGTTAAGCACTAACCATTTCCGGCTGACATCCCGTCCCGAGTGGGCT 551

QY 128 TATATCAGTATCACATTTGACTATAACCCACTGATGAAGCCAGAGACTCCGTTTCAGTCT 187
 DB 552 TATATCAGTATCACATTTGACTATAACCCACTGATGAAGCCAGAGACTCCGTTTCAGTCT 611

QY 188 TTCCTTTTCAACACAGAGATCTAANTGGAAGTGTCTGTTTGAATGAAGCATATTAT 247
 DB 612 TTCCTTTTCAACACAGAGATCTAANTGGAAGTGTCTGTTTGAATGAAGCATATTAT 671

QY 248 TTTTACCTAAAGAGCTACAGCAAAAGGTTTACTGAAGTTTATAGTAAGCCCGGAATGGAG 307
 DB 672 TTTTACCTAAAGAGCTACAGCAAAAGGTTTACTGAAGTTTATAGTAAGCCCGGAATGGAG 731

QY 308 AGGATGTGAGGATAACGATCACTTTAAACAAATGAACTTCCACTATACACCAACTTGT 367
 DB 732 AGGATGTGAGGATAACGATCACTTTAAACAAATGAACTTCCACTATACACCAACTTGT 791

QY 368 TGCAGTCTATATATATTTTCAGAGGCTTTTGAAGATCATGAATTTGCAACAAATG 427
 DB 792 TGCAGTCTATATATATTTTCAGAGGCTTTTGAAGATCATGAATTTGCAACAAATG 851

QY 428 GACGAAATATTATTAACCCAAATGACCCAAATGATATTCRAAGTCACAGGTTGGTGATTT 487
 DB 852 GACGAAATATTATTAACCCAAATGACCCAAATGATATTCRAAGTCACAGGTTGGTGATTT 911

QY 488 GGCCTGGCTTCACTACTTCCATCTTCAGTATGAAGAACAGCATCATGCTCTGCACTGAG 547
 DB 912 GGCCTGGCTTCACTACTTCCATCTTCAGTATGAAGAACAGCATCATGCTCTGCACTGAG 971

QY 548 TTAGCCATAAAGTCTTCCAGTGAAGCTGTTTGGATTTTCATGTTCACTTTTATATATC 607
 DB 972 TTAGCCATAAAGTCTTCCAGTGAAGCTGTTTGGATTTTCATGTTCACTTTTATATC 1031

QY 608 AGACAGAGAAATATAAATTTCAAGACAGGTTTCCAAAGACATTAATAGTTAGTTGTTTC 667
 DB 1032 AGACAGAGAAATATAAATTTCAAGACAGGTTTCCAAAGACATTAATAGTTAGTTGTTTC 1091

QY 668 TTACCAAGTATTAACATTAAGACATACAGAGTGGATGATTTGACTGGGACAGAAATCCCA 727
 DB 1092 TTACCAAGTATTAACATTAAGACATACAGAGTGGATGATTTGACTGGGACAGAAATCCCA 1151

QY 728 AGAGCACTTTAAAGAAAGCCGAGCCGCTCTGAAGTCAGCTTCTTAGAATACATACAGGAAGC 787
 DB 1152 AGAGCACTTTAAAGAAAGCCGAGCCGCTCTGAAGTCAGCTTCTTAGAATACATACAGGAAGC 1211

QY 788 AATACACCAAGAGATCACCGACTTGAAGCAGCCTGTCTTGGTCAGCCAGCCCAAGAGAA 847
 DB 1212 AATACACCAAGAGATCACCGACTTGAAGCAGCCTGTCTTGGTCAGCCAGCCCAAGAGAA 1271

QY 848 GGCCTGGCTTGGGGGACACTGTCAGGGCTGCATGCTCATTTCTGAGCTCTGCTATC 907
 DB 1272 GGCCTGGCTTGGGGGACACTGTCAGGGCTGCATGCTCATTTCTGAGCTCTGCTATC 1331

QY 908 TTAAGCTCTAACCTGATAAATGCTAATGATTTTAAAGCTGATGAAAGCTTAGCCGTTTC 967
 DB 1332 TTAAGCTCTAACCTGATAAATGCTAATGATTTTAAAGCTGATGAAAGCTTAGCCGTTTC 1391

QY 968 ATACAAGACTTAACCTCCAGAGCAAAAGGACGCTGAAGTGGGACGACTCATTTGATACATTC 1027
 DB 1392 ATACAAGACTTAACCTCCAGAGCAAAAGGACGCTGAAGTGGGACGACTCATTTGATACATTC 1451

QY 1028 ATAAAAACGATTAATGTTTCAAGGAGGCTTCCAGACTGGGGTTTTCAGCTTTTCATTTCCAACT 1087
 DB 1452 ATAAAAACGATTAATGTTTCAAGGAGGCTTCCAGACTGGGGTTTTCAGCTTTTCATTTCCAACT 1511

QY 1088 TACTGTCTTCTCAGGAAGAAATTTTGCACACAGAAAAAGATTCAACAAGGTGGAAAAACAT 1147
 DB 1512 TACTGTCTTCTCAGGAAGAAATTTTGCACACAGAAAAAGATTCAACAAGGTGGAAAAACAT 1571

QY 1148 TTGATTAACAATCCCAATTTTGCAGATTTGCTCCAAAGAAACAAAGAGGTGCACATTAATTA 1207
 DB 1572 TTGATTAACAATCCCAATTTTGCAGATTTGCTCCAAAGAAACAAAGAGGTGCACATTAATTA 1631

QY 1208 GTGTTAAAGCCACTAGATAAATCTGTTTGAATCTATACGCAAGAAATATTAAGCAGAGCA 1267
 DB 1632 GTGTTAAAGCCACTAGATAAATCTGTTTGAATCTATACGCAAGAAATATTAAGCAGAGCA 1691

QY 1268 ATTCAATTGATACAAAAATCTATTTAAAGTTACACAGCAATGGGCATGCAAAATGAGAAAAG 1327
 DB 1692 ATTCAATTGATACAAAAATCTATTTAAAGTTACACAGCAATGGGCATGCAAAATGAGAAAAG 1751

QY 1328 CAATAATGATTGAAGTGGATGACAGAACTGAAGCTTACTTAAGAGTCTTACAGCAAAAGG 1387
 DB 1752 CAATAATGATTGAAGTGGATGACAGAACTGAAGCTTACTTAAGAGTCTTACAGCAAAAGG 1811

QY 1388 TCACAGCAGACACCCAGATAGTTGTTCTGTCTGTCTGTTTGTCAAGTAATCGGAAGACAAATACG 1447
 DB 1812 TCACAGCAGACACCCAGATAGTTGTTCTGTCTGTCTGTTTGTCAAGTAATCGGAAGACAAATACG 1871

QY 1448 ATGCTATTAATAAATACCTGTGTGACAGATTGCCCTACCCCAAGTCAGTGTGTGGTGGCC 1507
 DB 1872 ATGCTATTAATAAATACCTGTGTGACAGATTGCCCTACCCCAAGTCAGTGTGTGGTGGCC 1931

QY 1508 GAACCTTAGGCAAAACAGCAAACTGTCTAGGCAATGCTTACAAAGATTGCCCTACAGATGA 1567
 DB 1932 GAACCTTAGGCAAAACAGCAAACTGTCTAGGCAATGCTTACAAAGATTGCCCTACAGATGA 1991

QY 1568 ACTCAAGATGGGAGGAGAGCTCTGAGGGTGGACATCCCCCTGAAGCTCGTGATGATCG 1627
 DB 2051 ACTCAAGATGGGAGGAGAGCTCTGAGGGTGGACATCCCCCTGAAGCTCGTGATGATCG 2051

QY 1628 TTGCAATCGATTGTTTACATGACATGACAGCTGGGGGAGGTCATCGAGAGATTTGTTG 1687
 DB 2052 TTGCAATCGATTGTTTACATGACATGACAGCTGGGGGAGGTCATCGAGAGATTTGTTG 2111

QY 1688 CCAGCATCAATGAAGGATGACCCGCTGTTTCTCACTGATATTTTCAGGATGAGGAC 1747
 DB 2112 CCAGCATCAATGAAGGATGACCCGCTGTTTCTCACTGATATTTTCAGGATGAGGAC 2171

QY 1748 AGGAGCTGGTAGATGGGCTCAAGCTCTGCCTGCAAGCGGCTCTCAGGCTTTGGAATAGCT 1807
 DB 2172 AGGAGCTGGTAGATGGGCTCAAGCTCTGCCTGCAAGCGGCTCTCAGGCTTTGGAATAGCT 2231

Db 1177 AAGAGCACCTTCAGAGAGGGGATGGCTCGAGAGTCACTCTCTGGAGTACTACAGGAA 1236
Qy 786 GCAATACAAACAGAGATCACCGACTTGAAGCAGCCTCTCTTGGTCAAGCCAGCCCAAGAG 845
Db 1237 GCAATACAAACAGAGATCACCGACTTGAAGCAGCCTCTCTTGGTCAAGCCCAAGAG 1296
Qy 846 AAGCGGGGCGCTGGGGGACACTCCAGGGCTGCCATGCTCATCTGAGCTCTGCTA 905
Db 1297 GAGGAGAGCGCGCGGCGACCTCGCTGGCGAGCTATGNTCATCCCTGAACTCTGCTA 1356
Qy 906 TCTTACAGTCTAAGCTGATTAATTCGTAATTCATTTAACGTGATGAAGACTTAGCCGT 965
Db 1357 TCTCACAGCCTGACTGATTAATTCGTAATTCATTTAACGTGATGAAGACTTAGCCGT 1416
Qy 966 TCATACAGACTAATTCACAGACCAAGGAGCGGTGAAGTGGGAGCAGCTCATTTGATAT 1025
Db 1417 GCACAGCGGCTGACCCCTGAGCAGCGCAGCGGAGGTGGCCGCCCTCATCGACTACAT 1476
Qy 1026 TCATAAAACGATATGTTCAAGGGAGCTTCGAGACTTCGAGCTGGGTTTGAAGTTGATTC 1085
Db 1477 CCACAAGGATGACAAATGTCAGAGAGAGCTTCGAGACTGGGCGCTGAGCTTCGACTCAA 1536
Qy 1086 CTTACTGCTCTCTCAGAGAGATTTTCCAAACAGAAAGATTCCACCAAGTGGAAAC 1145
Db 1537 CTTGCTGCTCTCTCGAAGAAATCTTCAATCTGAGAAGATTCACCAAGGCGGAAAGAC 1596
Qy 1146 ATTTGATTTACAAATTCACAAATTTGCAAGATTTGTCACAAAGAAAGATTCCACCAAGTGGAAAC 1205
Db 1597 GTTTGATTTACAAATTTGCAAGATTTGTCACAAAGAAAGATTCCACCAAGTGGAAAC 1656
Qy 1206 TAGTGTAAAGCACTAGATAATCTGCTGTGATCTATACGGAAGAAATATGAAGCAGC 1265
Db 1657 CAGCGTGAAGCAATGGGATACTGCTGCTGATCTATACCGCAGCAATTTATGAAGCAGC 1716
Qy 1266 CAATTCATTTGATCAAAATCTATTTAAAGTTTACACAGCCATGGGCATGCAATGAGAA 1325
Db 1717 CAATTCATTTGATCAAAATCTGTTTCAAGAGTACTCCAGCCATGGGCATGCAATGAGAA 1776
Qy 1326 ASCAATATGATTTGAAGTGGATGACAGAACTGAAGCCTACTTAAGATCTTTACAGCAAA 1385
Db 1777 GGCATCATGATCGAGTGGATGACAGAAACAGAACTTATCTGAGAGCCTTCAGAGAA 1836
Qy 1386 GGTACAGCAGACACCCAGATGTTGCTGTCTGTGTTGTTGTTGTTGTTGTTGTTGTTGTT 1445
Db 1837 GGTGACGTGACAGACTCAGATAGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTT 1896
Qy 1446 CGATGCTATTAAAAATACCTGTGTACAGATTGCGCTACCCCAAGTCAAGTGTGTGTGTCG 1505
Db 1897 TGATGCCATCAGAGTACTTGTGTACAGACTGCCCCCAAGTCAAGTGTGTGTGTTGTTGTTGTT 1956
Qy 1506 CCGAACCTTAGCAACCAAGCAATGTTGATGCCCATTTGTTGTTGTTGTTGTTGTTGTTGTTGTT 1565
Db 1957 CCGGACCTGGGAGCAGCAAAAGTCAATGSCCATTTGCCCAAGATCGCCCTGCAGAT 2016
Qy 1566 GAACTGCAAGTGGAGAGAGCTCTGAGAGTGGACATCCCTTGAAGTCTGTGATGAT 1625
Db 2017 GAACTGCAAGTGGAGAGAGCTCTGAGAGTGGACATCCCTTGAAGTCTGTGATGAT 2076
Qy 1626 CGTTGGCATCGATTGTTACCATGACATGACAGCTGGGCGAGGTCAATCGCAGAGTTTGT 1685
Db 2077 CGTTGGCATCGATTGTTACCATGACATGACAGCTGGGCGAGGTCCATCGCAGAGTTTGT 2136
Qy 1686 TGCCAGCATCAATGAAGGATGACCCGCTGTTCTCAGCTGCAATTTTCAGGATAGAGG 1745
Db 2137 CGCCAGCATCAATGAAGGATGACCCGCTGTTCTCCCGCTGCTCTTTTCAGGACCGCGG 2196
Qy 1746 ACAGAGCTGTTAGATGGCTCAAGTCTGCTGCAAGGGCTCTGAGGGCTTGAATAG 1805
Db 2197 GCAGAGCTGTTGATGTTCTCAAGTGTGTTGCAAGTGTCTTGAAGTGTCTTGAAGGCTTGAAGTGG 2256
Qy 1806 CTGCAATGATACATGCGCCAGCGGATCATCTGTACCGGATGCGGTAGGAGCGGCA 1865
Db 2257 CTGCAATGATACATGCGCCAGCGGATCATCTGTACCGGATGCGGTAGGAGCGGCA 2316

Qy 1866 GCTGAAACACTGGTGAACCTACGAAGTCCACAGTCTTTTGGATTGTCTAAATCCATTGG 1925
Db 2317 GCTGAAGACCTGGTCAATTTATGAGTCCACAGTTCCTAGATTGCTCAAGTCAGTCGG 2376
Qy 1926 TAGAGGTTACACCCCTAGACTAAGGTAATTTGGTGAAGAAAGAGTGAACACAGATT 1985
Db 2377 GAGAGGTTACAAACCCAAAGACTGACTGTAATCGTGGTGAAGAAAGCGTGTCAATGCCAGTT 2436
Qy 1986 TTTTGTCTCAGTCTGGGAGAGACTTTCAGAAATCCACTTCTCTGGACAGTTATTGATGTAGA 2045
Db 2437 TTTTGTCTCAGTCTGGGAGAGACTTTCAGAAATCCACTTCTCTGGACAGTTATTGATGTAGA 2496
Qy 2046 GGTTCACAGACCAAGATGTTGATGACTTTTATCGTCAGCAGGCTGTGAGAACTGGTAG 2105
Db 2497 AGTCACAGACCAAGATGTTGATGACTTTTATCGTCAGCAGGCTGTGAGAACTGGTAG 2556
Qy 2106 TGTTCCTCCACACATTACAAATGTCATATGACAAACAGCGGCTGAGGCCAGACACAT 2165
Db 2557 TGTTCCTCCACACATTACAAATGTCATATGACAAACAGCGGCTGAGGCCAGACACAT 2616
Qy 2166 ACAGCGCTTCGACCTACAAAGCTGTGCCACATCTATTCAACTGGCCAGGTGTCATTGCTGT 2225
Db 2617 CCAGCGCTGACATACAAAGTCTGCCACGTTACTTAAATTTGGCTGGAGTCAATCCAGT 2676
Qy 2226 TCCTGCTCTCTTGGCAGTACGCCACAGCTGGCTTTTCTTTGTTGGCCAGAGTATTCACAG 2285
Db 2677 CCCTGCACTTGGCAGTATGCACACAGCTGGCTTCTCTGTTGGCCAGAGATTCACAG 2736
Qy 2286 AGAGCCAAATCTGTCATCTGCAAAACCGCTTTTACTACTCTTAA 2328
Db 2737 AGAGCCAAATCTGTCATCTGCAAAACCGCTTCTCTGTTGGCCAGAGTATTCACAG 2799

RESULT 3

ABV89287

ID ABV89287 standard; cDNA; 2292 BP.

XX ABV89287;

AC XX (first entry)

DT 13-DEC-2002

XX Human colon cancer related cDNA SEQ ID NO 2602.

XX Human; colon; cancer; cytostatic; tumour; gene therapy; vaccine; gene;

XX ss.

XX Homo sapiens.

OS WO200258534-A2.

XX 01-AUG-2002.

XX 19-NOV-2001; 2001WO-US43704.

XX 20-NOV-2000; 2000US-252222P.

XX 06-FEB-2001; 2001US-267011P.

XX 28-MAR-2001; 2001US-279670P.

XX 10-JUL-2001; 2001US-304037P.

XX (CORI-) CORIXA CORP.

XX Stolk JA, Xu J, Chenault RA, Meagher MJ, Secretist H, King GE;

XX WPI; 2002-608400/65.

XX P-PSDB; ABP67995.

XX New isolated tumor colon polynucleotide and polypeptide, useful for the diagnosis, prevention and/or treatment of cancer, in particular colon cancer

XX Claim 1; SEQ ID NO 2602; 266pp + Sequence Listing; English.

CC The invention relates to a human colon tumour expressed polynucleotide
CC (I) encoding a polypeptide (II, ABP67991-ABP67996) comprising: (i) any of
CC 2600 fully defined nucleotide sequences (ABV8689-ABV8289); (ii)
CC complements of (i); (iii) at least 20 contiguous residues of (i); (iv)
CC sequences that hybridize to (i), under moderately stringent conditions;
CC (v) sequences having at least 75% or 90% identity to (i), or (vi)
CC degenerate variants of (i). The compositions and methods of the present
CC invention are useful for the diagnosis, prevention and/or treatment of
CC cancer, particularly colon cancer. (i) can be used in gene therapy and
CC (ii) and (iii) are useful in pharmaceutical compositions such as vaccines.
CC Note: The sequence data for this patent did not form part of the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published_pct_sequences.

XX Sequence 2292 BP; 692 A; 435 C; 512 G; 653 T; 0 other;

Query Match 67.4%; Score 1569; DB 24; Length 2292;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 1569; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY	760	GTCAGCTTCTTAGAATACACAGGAGCAATACACCAAGAGATCACCGACTTGAGCAG	819
DB	4	GTCAGCTTCTTAGAATACACAGGAGCAATACACCAAGAGATCACCGACTTGAGCAG	63
QY	820	CCTGTCTTGGTCAGCCAGCCCAAGAGAGGGGGCCCTGGGGGACACTGCCAGGGCCT	879
DB	64	CCTGTCTTGGTCAGCCAGCCCAAGAGAGGGGGCCCTGGGGGACACTGCCAGGGCCT	123
QY	880	GCCATGCTCATCTCTGAGCTCTGATCTTACAGCTCTAATCTGATTAATAATGCTAATGAT	939
DB	124	GCCATGCTCATCTCTGAGCTCTGATCTTACAGCTCTAATCTGATTAATAATGCTAATGAT	183
QY	940	TTTAACTGATGAAAGACTTACGCTTTCATACAGACTTAACTCCAGAGCAAGGACAGCT	999
DB	184	TTTAACTGATGAAAGACTTACGCTTTCATACAGACTTAACTCCAGAGCAAGGACAGCT	243
QY	1000	GAACTGGGACGACTCATTTGATTTACATTTTCAAAAAAGTAACTTTCAAGGGAGCTTGA	1059
DB	244	GAACTGGGACGACTCATTTGATTTACATTTTCAAAAAAGTAACTTTCAAGGGAGCTTGA	303
QY	1060	GACTGGGTTTGAAGCTTTGATTTCAACTTACTGCTCTCTCAGGAGAAATTTTGCACACA	1119
DB	304	GACTGGGTTTGAAGCTTTGATTTCAACTTACTGCTCTCTCAGGAGAAATTTTGCACACA	363
QY	1120	GAAAGATTCACCAAGTGGGAAACATTTGATTTACATTTCAACTTCCAGAGCAAGGACAGCT	1179
DB	364	GAAAGATTCACCAAGTGGGAAACATTTGATTTACATTTCAACTTCCAGAGCAAGGACAGCT	423
QY	1180	AAAGAAACAAAGAGTGCACCATTAATTTAGTTTAAAGCCACTAGATAAAGCTGTTGATC	1239
DB	424	AAAGAAACAAAGAGTGCACCATTAATTTAGTTTAAAGCCACTAGATAAAGCTGTTGATC	483
QY	1240	TATACGGAAGAAATATGAGAGCCCAATTTCAATTTGATTAACAAATCTATTAAAGTTACA	1299
DB	484	TATACGGAAGAAATATGAGAGCCCAATTTCAATTTGATTAACAAATCTATTAAAGTTACA	543
QY	1300	CCAGCCATGGGCAATCAATGAGAAAGCAATTAATTTGAGTGTGATGATGATGATGATGAT	1359
DB	544	CCAGCCATGGGCAATCAATGAGAAAGCAATTAATTTGAGTGTGATGATGATGATGATGAT	603
QY	1360	GCTTACTTAAAGAGTCTTACAGCAAAAGGTTCACAGCAGACACCCAGATAGTTGCTGTCTG	1419
DB	604	GCTTACTTAAAGAGTCTTACAGCAAAAGGTTCACAGCAGACACCCAGATAGTTGCTGTCTG	663
QY	1420	TTGTCAAGTAATTCGAAGCAAAATACGATGCTATTAAATAATACCTGTGTACAGATTGC	1479
DB	664	TTGTCAAGTAATTCGAAGCAAAATACGATGCTATTAAATAATACCTGTGTACAGATTGC	723
QY	1480	CCTACCCCAAGTCAGTGTGTGGCCGCAACTTAGGCAACACAGCAAACTGTCTGTGCC	1539
DB	724	CCTACCCCAAGTCAGTGTGTGGCCGCAACTTAGGCAACACAGCAAACTGTCTGTGCC	783
QY	1540	ATTGCTACAAAGATTGCCCTACAGATGAATGCAAGATGGGAGAGAGCTCTGGAGGGTG	1599

DB	784	ATTGCTACAAAGATTGCCCTACAGATGAATGCAAGATGGAGAGAGCTCTGGAGGGTG	843
QY	1600	GACATCCCTCGAAGCTCTGATGATCTGGGCAATGATTTTACCATGATGATGATGATGAT	1659
DB	844	GACATCCCTCGAAGCTCTGATGATCTGGGCAATGATTTTACCATGATGATGATGATGAT	903
QY	1660	GGGGGAGGTCAATCGCAGGATTTGTTCCAGCATCAATGAAGGATGACCCGCTGGTTTC	1719
DB	904	GGGGGAGGTCAATCGCAGGATTTGTTCCAGCATCAATGAAGGATGACCCGCTGGTTTC	963
QY	1720	TCACGCTGCATATTTCAGGATAGAGGACAGAGCTGCTAGATGGGCTCAAAAGTCTGGCTG	1779
DB	964	TCACGCTGCATATTTCAGGATAGAGGACAGAGCTGCTAGATGGGCTCAAAAGTCTGGCTG	1023
QY	1780	CAAGCGCTCTGAGGGCTTGGATAGCTGCAATAGTACATGCCAGCCGATCATCTGTG	1839
DB	1024	CAAGCGCTCTGAGGGCTTGGATAGCTGCAATAGTACATGCCAGCCGATCATCTGTG	1083
QY	1840	TACCGCATGGCTGAGGAGCGGCGCAGCTGAAACACTGGTGAACCTAGAACTGCAAGTGCACAG	1899
DB	1084	TACCGCATGGCTGAGGAGCGGCGCAGCTGAAACACTGGTGAACCTAGAACTGCAAGTGCACAG	1143
QY	1900	TTTTTGGATTGCTTAAATCCATTTGTTAGAGTTTACAACTAGACTAACCGTAACTTGTG	1959
DB	1144	TTTTTGGATTGCTTAAATCCATTTGTTAGAGTTTACAACTAGACTAACCGTAACTTGTG	1203
QY	1960	GTGAAGAAAGAGTGAACACCAAGATTTTGTCTCAGTCTGAGGAGAGACTTCAGATGCA	2019
DB	1204	GTGAAGAAAGAGTGAACACCAAGATTTTGTCTCAGTCTGAGGAGAGACTTCAGATGCA	1263
QY	2020	CTTCTCTGAAACAGTATTGATGATGATGATGATGATGATGATGATGATGATGATGATGAT	2079
DB	1264	CTTCTCTGAAACAGTATTGATGATGATGATGATGATGATGATGATGATGATGATGATGAT	1323
QY	2080	GTGAGCCAGGCTGTGAGAGTGTGATGTTTCTCCACACATTAATGATCATCTATGAC	2139
DB	1324	GTGAGCCAGGCTGTGAGAGTGTGATGTTTCTCCACACATTAATGATCATCTATGAC	1383
QY	2140	AACAGCGGCTTGAAGCAGACACCATACAGCGCTTGACCTACAGCTGTGCCATCTAT	2199
DB	1384	AACAGCGGCTTGAAGCAGACACCATACAGCGCTTGACCTACAGCTGTGCCATCTAT	1443
QY	2200	TACAACTGGCCAGGTGTCATTCGTTCTCTCTCCAGTACGCCACCAAGCTGCT	2259
DB	1444	TACAACTGGCCAGGTGTCATTCGTTCTCTCTCCAGTACGCCACCAAGCTGCT	1503
QY	2260	TTTCTTGTGGCCAGAGTATTACAGAGAGCCAAATCTGTCACTGTCAAAACCGCTTTAC	2319
DB	1504	TTTCTTGTGGCCAGAGTATTACAGAGAGCCAAATCTGTCACTGTCAAAACCGCTTTAC	1563
QY	2320	TACCTCTAA 2328	
DB	1564	TACCTCTAA 1572	

RESULT 4
ABQ61109
ID ABQ61109 standard; cDNA; 2009 BP.
XX AC ABQ61109;
XX XX
DT 26-FEB-2003 (first entry)
XX
DE Human mRNA sequence encoding sequence.
XX
KW Neuroprotective; immunomodulator; cancer;
KW Cytostatic; anti-inflammatory; gene therapy; nutritional supplement;
KW wound; burn; ulcer; Alzheimer's disease; Huntington's disease;
KW amyotrophic lateral sclerosis; autoimmune disorder; inflammation;
KW vulnary; gene; ss.
OS Homo sapiens.

XX WO200231111-A2.
 PN
 XX
 XX PD 18-APR-2002.
 XX
 XX 11-OCT-2001; 2001WO-US27760.
 XX
 XX 12-OCT-2000; 2000US-0687527.
 PR (HYSE-) HYSEQ INC.
 XX
 XX Xue YT, Liu C, Zhou P, Asundi V, Zhang J, Zhao QA, Ren F;
 PI Tang AJ, Yang Y, Wehrman T, Drmanac RT;
 XX
 XX WPI; 2002-426278/45.
 DR N-ESDB; ABP43865.
 XX
 XX New polypeptides and their encoded proteins, useful as nutritional
 PT sources or supplements, or in gene therapy, particularly for treating
 PT wounds, Alzheimer's disease, amyotrophic lateral sclerosis, cancer or
 PT inflammation -
 XX
 XX Claim 1; SEQ ID # 322; 357pp + sequence listing; English.
 XX
 XX The invention relates to 446 newly isolated polynucleotide sequences.
 CC The activity of polynucleotides of the invention may be described as,
 CC vulnery, neuroprotective, immunomodulator, cytostatic and
 CC anti-inflammatory. Compositions comprising nucleic acids of the invention
 CC are useful for treating a mammalian subject, or as nutritional sources or
 CC supplements. These are useful in gene therapy, particularly for treating
 CC wounds, burns or ulcers, Alzheimer's disease, Huntington's disease,
 CC amyotrophic lateral sclerosis, autoimmune disorders, cancer or
 CC inflammation. The nucleic acids and polypeptides are also useful in
 CC diagnostic and research methods. The sequences given in records
 CC AB060788-AB061233 represent polynucleotides of the invention.
 CC NOTE: The sequence data for this patent did not form part of the printed
 CC specification, but was obtained in electronic format directly from WIPO
 CC at ftp.wipo.int/pub/published_pat_sequences.
 XX
 XX Sequence 2009 BP; 595 A; 411 C; 458 G; 545 T; 0 other;
 SQ
 Query Match 19.7%; Score 458.4; DB 24; Length 2009;
 Best Local Similarity 57.2%; Pred. No. 6.9e-126;
 Matches 894; Conservative 0; Mismatches 661; Indels 9; Gaps 3;
 QY 765 CTTCTTGAATACTACAGGAAGCAATACAAACAGAGATCCCGACTTGAAGCAGCCTGT 824
 DB 15 CTAGGTGATTACTACAAGCAGCAGTATGATATTACTGTATCGGACCTGTAATCAGCCCAT 74
 QY 825 CTTGTCTCAGCCCAAGAGAGCGGGCCCTGGGGGGACATGCCAGGGCCCTGCCAT 884
 DB 75 CTTGTGTAGTCTGTTAAG---AAGAGAGAAATGACACAGTGAAGCTCAGCTCGGCCA 131
 QY 885 GCTCATCTCCTGAGCTCTCTATCTTACAGGTCTTAAGTATGATGATGATTTAA 944
 DB 132 CCTGATACCTGAGCTCTGCTTTCTTAACAGGCTGACTGACAGGCAACATCTGATTTCCA 191
 QY 945 CGTGAATGAAGACTTAGCCGTTTCATACAGACTAATCTCCAGAGCAAGGCGAGCTGAAGT 1004
 DB 192 GCTGATGAAGGCTGTGGGTGAAAGACACAGCTCTCAGTCTCTTCAGGCCGCGCAGCGCCT 251
 QY 1005 GGGAGCACTCATGATTATCATTTATTAACATTAACAGATATCTCAAGGGAGCTTCGAGACTG 1064
 DB 252 GCGCAGGCTTTGGACACATCCAGAGATACCAATGCTCGCTTGAACATGAGAGACCTG 311
 QY 1065 GGGTTTGAAGCTTTGATTCACACTTACTCTCTCTCAGAGAAATTTTGCAGAAAGAAA 1124
 DB 312 GGGACTGATTTTGAAGCCAGATA---TCTCTGACTGGCGGATGTGCTTTCAGAAAA 368
 QY 1125 GATTCCAAAGTGGAAAAAATTTGATTAACAATCCAAATTTGAGATTTGGTCCAAAGA 1184
 DB 369 AATATTATGCAAGACCCACATAT---GTCAACCTGTGTCTGCTGACTGCTCAGGA 425

QY 1185 AACAGAGGTGCACCAATTAATAGTGTTAAGCCACTAGATAAAGTGGCTGTTGATCTATAC 1244
 DB 426 TATTTCGAACCTTGCAGAGATTTTAAATGCACAGTCTTTTGAATACCTGGTTGATTTATGTAG 485
 QY 1245 GCGAAGAAATATGAAGCAGCCCAATTCATTGATACAAAAATCTATTTAAAGTTACACGAGC 1304
 DB 486 CGACAGAACTGAATATATTTGCGGAGAGCTTTCTGAATCTGCTTGAGAAGAGTTGAGGTTT 545
 QY 1305 CATGGGCATGCAATGAGAAAAAGCAATATGATGAGTGAAGTGGATGACAGAACTGAAGCCTA 1364
 DB 546 CATGGGATTTAACTGTGACCTACCCCAAATCATAAAGTACAGAAATCCAGCTGCATT 605
 QY 1365 CTTAAGAGCTTTACAGCAAAAGTTCACAGCAGACACCCAGATAGTTGCTGTCTGTCTGTC 1424
 DB 606 TGTTAGAGCTTATACAGCAATATGTGTATCTCTGATGTTTCACTGCTGTAATGTGCACTTCGC 665
 QY 1425 AAGTAATCGAAGGACAAATACGATGCTATTAATAAATACCTGTGTACAGATTCGCCCTAC 1484
 DB 566 TTCTAATCAGAGACCTATTATGATTCATTAATAAATATTTGAGCTCAGATGCCAGT 725
 QY 1485 CCCAAGTCAGTGTGTGGTGGCCGAACTTTAGGCAACACAGCAAACTGTCTATGGCCATTGC 1544
 DB 726 CCCAAGCCATGTGTGCTTCTCGACCTTGAATAACAGGGCATGATGATGATGATCGC 785
 QY 1545 TACAAGATTGCCCTACAGATGAATCTCAAGATGGGAGGAGAGCTCTGGAGGGTGGACAT 1604
 DB 786 CACCAAGATCGCTATGAGATGACTTTCAAGCTCGGAGGCGAGCTGTGGGTGTGGAAAT 845
 QY 1605 CCCCTCAAGCTCGTGTATGATGCTGTGCATCGATGTTTACCATGACATGACAGCTGGGCG 1664
 DB 846 ACCTTAAATCCCTGATGCTGATGCTGCTGATGCTGTAAGATGCATCAGCAAGGA 905
 QY 1665 GAGTCTCAATCGAGGATTTTGTCCAGCATCAATGAAGGATGACCCGCTGGTTCTCACG 1724
 DB 906 CGTGATGTTGTTGGATGCTGGCGCTGTTAAACCCAGAAATCACAGGTGGTTTCCCG 965
 QY 1725 CTGCATATTTCAGATFAGAGGACAGGAGCTGTGATGGGCTCAAAGTCTGCCTGCAAGC 1784
 DB 966 CTGTATCTTTCAGAGAAACAATGATGATGTTGTCAGATTTGCTTGAAGCTTTTCATGACTGG 1025
 QY 1785 GGCTCTGAGGCTTTGGAATAGCTGCAATGATGATGCCAGCCGAGATCATCGTGTACCG 1844
 DB 1026 AGCATCAACAATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1085
 QY 1845 CGATGGCTTAGGAGACGGCCAGCTGAAAAACACTGGTGAACACTAGCAAGTGCACAGTTTTT 1904
 DB 1086 TGCTGTGTAGGGGATGCTGAGCTGAAAAACACTTATTGATATGAACTCCACAGCTGCT 1145
 QY 1905 GGATTGCTAAATCCATTGGTAGAGGTTTACACCTTAGACTAACCGTAAATGTTGGTGAA 1964
 DB 1146 GAGCAGTGTGGCAGAAATCCAGCTCAAAATACAGCTCAAGACTGTCGGTGATTTGGTCAAG 1205
 QY 1965 GAAAAGAGTGAACACACAGATTTTGTCTGCTGAGGAGAACTTCAGAAATCCACTTCC 2024
 DB 1206 GAAGAGTGCATGCCAGATTTCTTACCGAATGAACCGCACTGTACAGAACCCCACT 1265
 QY 2025 TGAACAGTATTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 2084
 DB 1266 TGGCACTGTTGTGGATTCAGAAAGCAACCGTAAACGAATGGTATGACTTTTATCTGATCAG 1325
 QY 2085 CCAGGCTGTGAGAGTGTGATGTTTCTCCACACATTAACAATGCTCATCTATGACACAG 2144
 DB 1326 CCAGGTGGCTTCCGGGAACTGTAGTCTTACTCTACTACTACTACTACTACTACTACTACTACT 1385
 QY 2145 CGGCTTGAAGCCAGACACACATACAGCGCTTGAACCTGAGCTGTGCCACATCTATTACAA 2204
 DB 1386 CGCTTGAAGCCAGACATATGAGAGACTTATTAATTAATTTGCGCACTCTGATCTACAA 1445
 QY 2205 CTGSCCAGGTGTCTATGCTGCTTCT 2264
 DB 1446 CTGSCCGGCGATAGTCTAGTGTCCAGCAACCATGTCAGTATGCTCAGCAAGCTGACCTTTCT 1505
 QY 2265 TGTGGCCAGAGTATTTCACAGAGAGCCAAATCTGTCTACTGTCTCAACCGCCCTTTACTACCT 2324

Db 1506 GGGGGCACAAAGCATTTCATTAAGAACCCAGCTGCTGGATTAGCCCAACCATCTCTTCTACCT 1565
 QY 2325 CTAA 2328
 Db 1566 GTGA 1569

RESULT 5
 AAF21880
 ID AAF21880 standard; DNA; 423 BP.
 XX
 AC
 XX AAF21880;
 DT 27-MAR-2001 (first entry)
 XX
 DE Human breast and ovarian cancer associated antigen gene SEQ ID 267.
 XX
 KW Human; breast cancer; ovarian cancer; cytostatic; immunosuppressive;
 KW neutropic; neuroprotective; antiviral; antiallergic; hepatotropic;
 KW antidiabetic; antiinflammatory; antiulcer; vulnery; anticonvulsant;
 KW antibacterial; antifungal; antiparasitic; cardiant; immune disorder;
 KW Addison's disease; allergy; autoimmune haemolytic anaemia;
 KW autoimmune thyroiditis; diabetes mellitus; Crohn's disease;
 KW multiple sclerosis; rheumatoid arthritis; ulcerative colitis;
 KW cardiovascular disorder; wound healing; neurological disease; ds.
 XX
 OS Homo sapiens.
 XX
 FN WO200055173-A1.
 XX
 PD 21-SEP-2000.
 XX
 XX 08-MAR-2000; 2000WO-US05891.
 XX
 PR 12-MAR-1999; 99US-0124270.
 XX
 PA (HUMA-) HUMAN GENOME SCI INC.
 XX
 PI Rosen CA, Ruben SM;
 XX
 DR WPI; 2000-611515/58.
 DR P-PSDB; AAB58977.
 XX
 PT New human breast and ovarian cancer associated gene sequences and the
 PT polypeptides encoded by these genes, useful in the prevention,
 PT treatment and diagnosis of cancer, immune disorders, cardiovascular
 PT disorders and neurological diseases -
 XX
 PS Claim 1; Page 686; 1299pp; English.
 XX
 CC Sequences AAF21614 - AAF22031 represent DNA sequences encoding human
 CC proteins AAB58711 - AAB59128. The DNA and protein sequences are
 CC associated with breast and ovarian cancer. Included in the invention are
 CC sequences AAF2032 - AAF2040 and AAB59129 which are used in the
 CC isolation and characterisation of the DNA and protein sequences of the
 CC invention. The breast and ovarian cancer associated DNA, protein, agonist
 CC or antagonist sequences exhibit cytostatic; immunosuppressive;
 CC neutropic; neuroprotective; antiviral; antiallergic; hepatotropic;
 CC antidiabetic; antiinflammatory; antiulcer; vulnery; anticonvulsant;
 CC antibacterial; antifungal; antiparasitic and cardiant activity. The
 CC polynucleotide and protein sequences are used in the diagnosis of cancer,
 CC particularly breast and ovarian cancer. The nucleic acid sequences,
 CC proteins, agonists and antagonists may also be used in the diagnosis,
 CC prevention and treatment of immune disorders e.g. Addison's disease,
 CC allergies, autoimmune haemolytic anaemia, autoimmune thyroiditis,
 CC diabetes mellitus, Crohn's disease, multiple sclerosis, rheumatoid
 CC arthritis and ulcerative colitis; cardiovascular disorders such as
 CC myocardial ischaemias; wound healing; neurological diseases such as
 CC cerebral anoxia and epilepsy; and infectious diseases.
 XX
 SQ Sequence 423 BP; 104 A; 98 C; 129 G; 88 T; 4 other;

Query Match 17.0%; Score 395.4; DB 21; Length 423;
 Best Local Similarity 98.8%; Pred. No. 1.8e-107;
 Matches 404; Conservative 4; Mismatches 0; Indels 1; Gaps 1;

QY 1474 GATTGCCCTACCCCAAGTCAGTGTGTGGTGGCCCGAACCTTAGGCAACAGCAAACTGTG 1533
 Db 14 GATTGCCCTACCCCAAGTCAGTGTGTGGTGGCCCGAACCTTAGGCAACAGCAAACTGTG 73

QY 1534 ATGGCCATTGCTACAAAGATTGCCCTACAGATGAATGCAAGATGGGAGGAGACTCTGG 1593
 Db 74 ATGGCCATTGCTACAAAGATTGCCCTACAGATGAATGCAAGATGGGAGGAGACTCTGG 133

QY 1594 AGGGTGGACATCCCCCTGAAAGCTCGTGCATGATCGTTGGCATCGATTGTTACCATGACATG 1653
 Db 134 AGGGTGGACATCCCCCTGAAAGCTCGTGCATGATCGTTGGCATCGATTGTTACCATGACATG 193

QY 1654 ACAGCTGGGGGAGGTCATTCGAGGATTTGTCAGCATCAATGAAGGATGACCCGC 1713
 Db 194 ACAGCTGGGGGAGGTCATTCGAGGATTTGTCAGCATCAATGAAGGATGACCCGC 253

QY 1714 TGCTTCTCAGCTGTCATATTTTCAGGATAGAGGACAGGAGCTGCTAGATGGCTCAAAGTC 1773
 Db 254 TGCTTCTCAGCTGTCATATTTTCAGGATAGAGGACAGGAGCTGCTAGATGGCTCAAAGTC 313

QY 1774 TGCTTGAAGCGGCTCTGAGGGCTTGGAATAGCTGCAATGAGTACATGCCAGCCGATC 1833
 Db 314 TGCTTGAAGCGGCTCTGAGGGCTTGGAATAGCTGCAATGAGTACATGCCAGCCGATC 373

QY 1834 ATCGTGTACCCGATGGCTGAGAGACGGCCAGCTGAAACACTGGTGA 1882
 Db 374 ATCGTGTACCCGCTGAGGGCTTGGAATAGCTGCAATGAGTACATGCCAGCCGATC 421

RESULT 6
 AAS13630
 ID AAS13630 standard; cDNA; 3649 BP.
 XX
 AC AAS13630;
 XX
 DT 18-DEC-2001 (first entry)
 XX
 DE
 XX
 OS
 XX Mammalia.
 XX
 PN WO200166752-A2.
 XX
 PD 13-SEP-2001.
 XX
 PF 07-MAR-2001; 2001WO-US07371.
 XX
 PR 07-MAR-2000; 2000US-0187518.
 PR 12-JAN-2001; 2001US-0261557.
 XX
 XX (WHED) WHITEHEAD INST BIOMEDICAL RES.
 PA
 XX Wang PJ, Page DC;
 XX
 DR WPI; 2001-570774/64.
 DR P-PSDB; AAU07866.
 XX
 XX Novel reproduction-specific protein, useful for treating disorders of
 XX reduced sperm count, enhancing/increasing sperm count and/or sperm
 XX activity -
 XX
 PS Claim 1; Fig 15; 151pp; English.
 XX
 CC The present invention relates to the isolation of novel mammalian and
 CC human reproduction-specific proteins (AAU07859-AAU07899), and the

CC nucleic acids encoding them. The nucleic acids encoding
 CC reproductive-specific proteins are useful for diagnosing infertility
 CC which is a result of reduced sperm count, reduced sperm motility,
 CC malformed sperm or combinations of these. The sequences of the invention
 CC are useful as markers for spermatogonial cells, for identifying genes or
 CC proteins characteristic of male infertility, diagnosing or aiding in
 CC the diagnosis of infertility in men, and for contraception in which
 CC sperm production or sperm count is reduced or defective sperm is
 CC produced. Antibodies to reproductive-specific proteins are useful for
 CC determining the presence of these proteins in a sample obtained from a
 CC man being assessed for infertility, for identifying the expression of
 CC genes in particular cell type or particular developmental stage, for
 CC studies of spermatogenesis, and for immunofluorescence of germ cells or
 CC in Western blots for assessing the presence of the protein the antibody
 CC binds. The sequences of the invention are also useful for treating
 CC disorders of the invention and/or for increasing sperm count and/or
 CC sperm activity. The nucleic acids of the invention are useful in gene
 CC therapy. AAS13623-AAS13647 represent cDNA sequences encoding for the
 CC mammalian reproduction-specific proteins of the present invention.

XX
 SQ Sequence 3649 BP; 947 A; 841 C; 911 G; 950 T; 0 other;

Query Match 16.4%; Score 381; DB 22; Length 3649;
 Best Local Similarity 52.7%; Pred. No. 1.4e-102;
 Matches 975; Conservative 0; Mismatches 855; Indels 21; Gaps 6;

QY 490 CTTGGCTTCACTTCTCCATCTTCAGTATGAAACAGCATCATGCTCTGCACGTGAGTT 549
 DB 1 CTTGGCTATCGGCTAGTATATCCGAGGACAGACGGGGCTCTCTCTCTCTCGCTGATGTC 60

QY 550 AGCCATAAAGTCCTTGAAGTGAGACTGTTTGGATTCATGTTCAACTTTATCATCAG 609
 DB 61 TCTCTAAGGTCATTCGGAACGACTCTGCTGATGTCATGTCATCATCAGCAG 120

QY 610 ACAGAAGACATAAATTTCAAGAACAGATTTCCTCAAGAACTTAATAGTTTATGTTGTTCTT 669
 DB 121 AACAGGAGCAC---TTCAGGACGAGTGCAGCAAGCTTCTGGTTGGCAGCATTTGTCAIC 177

QY 670 ACCAAGTATACATATAGACATACAGAGTGATGATATGATCTGGACCAAGATCCCAAG 729
 DB 178 ACGGGCTACACAATCGTACCTACCGAATCATGATGTGAGTGCAGTGCAGACCCCTTAA 237

QY 730 AGCACTTTAAGAAAGCCGCGCTCTGAAGTGCAGCTCTCTTGGTGCAGCCAGCCCA---AGAGA 846
 DB 298 TATGGGATCACAGTCAAGGAAGATGACCCAGCGCTGCTGATCCACCGGCCAGTGAGAGA 357

QY 847 AGCGGGCCCTGGGGGACACTGCCAGGGCTGCCATGCTCATCTCTGAGCTCTGCTAT 906
 DB 358 CAGAAATACCATGCGATGTTGCTGAAGGGCGAGATCCTGCTGCTGCCCGAGCTCTCCTTC 417

QY 907 CTTACAGGCTTAACTGATAAATGCGTAATGATTTTAAAGTGTGATAAGACTTACGCGTT 966
 DB 418 ATGACGGGATCCCTGAGAAGATGAAAGAGCTTTCAGGGCCATGAAGACTTGAATCAG 477

QY 967 CATCAAGACTTAACTCAGAGCAAGGCGAGCGTGAAGTGGGACGACTCATTTGATTAATT 1026
 DB 478 CAGATTAACTTGAGCCCAAGCAGCACCCAGGTCCTTTGGAAATGCCCTGCTSCAGAAAT 537

QY 1027 CATAAACAGTAATGTTCAAGAGGAGTTCGAGACTGGGGTTTGGATTTGATTCGAAC 1086
 DB 538 TCACAAACAGGACAGCCAGCAATGAGTGAACCGCTGGGGGCTCAGTCTGATTAAGAT 597

QY 1087 TTAAGTCTCTCTCAGGAAGATTTTCAACAGAAAGATTTCAACAGTGGGAAACA 1146
 DB 598 GTCCACAGATTGAGTGCCTCTCCCAATGAGAGGATCACTTAAGAAACATTTCA 657

QY 1147 TTTGATTACATCAATTTGAGATTGGTTCCTCAAGAAACAGAGGTGACCACTTAATT 1206
 DB 658 TTTTGT---CACATCGAGGGCTGAATCTGGGTTAAGGAAGTGCACAGAGTGTCTTCATT 714

QY 1207 AGTGTAAAGCCACTAGATAAATCTGGCTGTTGATCTATACGGAAGAAATTAAGACGACC 1266
 DB 715 CTAATATTCCCTGATCTTCTGGCACTCTTTTATCCAAAGAGAGCAATGGACCAAGCC 774

QY 1267 AATTCATTGATACAAATCTATTAAAGTTACACAGCCCATGGCATGCAAAATGAGAAA 1326
 DB 775 AGAAGACTGGTTAACTGTTTGGAAAGATTGCCGGGCCCATGGCATGCCGACAAAGCCCC 834

QY 1327 GCATAATGATTTGAAGT---GGATGACAGAACTGAAGCCCTTACTTTAAGAGTCTTACAGCAA 1383
 DB 835 CCAGCCTGGCTTGGCTGAGGATGACCGAATAGAGACCTATATCAGGACCATTCAGTCC 894

QY 1384 AAGGTACACAGAC-----ACCCAGATAGTTGTCTGTCTGTGTCAAGTAATCGGAAG 1437
 DB 895 TTACTGGGAGTTGAGGGAAGATACAAATGGTGTGTGATCATCATGATGGGACACAGGTAT 954

QY 1438 GACAAATACGATGCTATTAAATAATACCTGTGTACAGATTGCCCTACCCCAAGTCAAGTGT 1497
 DB 955 GATCTCTATGGAGCCATCAAGAAAGTGTGTGGTGCAGTCCCGCAGTCCGCTCACAGGTC 1014

QY 1498 GTGGTGGCCCGAACTTAGGCAAAACAGAAACTGTCTATGCGCCATTGCTACAAAGATTGCC 1557
 DB 1015 ATCAATGTCCGACCATTTGTCAGCCACCACAGGCTTCGAGAGCTGCTCAGAAATTTTA 1074

QY 1558 CTAAGATGAACCTGCAAGATGGAGAGAGCTCTGAGAGGTGAGACATCCCCCTGAAAGCTC 1617
 DB 1075 CTTTCAAGTGAACCTGAAACTGGGTGTGAGCTCTGGGAGTGTATTTCCGCTGAAACAA 1134

QY 1618 GTGATGATCGTTGGCATCGATTGTTACCATGACATGACAGCTGGGGAGGTCAATCGCA 1677
 DB 1135 CTAATGGTGAATGGATGGATGTACCATGACCCAGCAGAGGCGATCGCTCTGTGGTC 1194

QY 1678 GGATTTGTTCCAGCATCAATGAAGGATGACCCCGTGGTCTCACGCTGCATATTTCAG 1737
 DB 1195 GGGTTCTGGCCAGCATATAATCTCACACTCACCAATGGTACTCGAGGGTGGTTCAG 1254

QY 1738 GATAGAGACAGAGCTGTGTAGATGGGCTCAAGTGTGCTGCAAGCGGCTCTGAGGGCT 1797
 DB 1255 ATGCCCATCAGGAGATTGTGACAGAGCTGCAAGCTCTGCTGGTGGTCTCTTGAAGAAAG 1314

QY 1798 TGGAAATGCTGCAATGATGATACATGCCAGCGGATCATCGTGTACCGCATGGCGTAGGA 1857
 DB 1315 TATTATGAGGTGAACCATTTGTCTCCAGAGAAAATTTGTGGTGTACCGAGATGGAGTGTCT 1374

QY 1858 GACGGCCAGCTGAAAAACACTGGTGAACCTACAAAGTGCACAGATTTTGGATTTGCTAAAA 1917
 DB 1375 GATGGCCAGCTAAAGACAGTTGCCACTACAGATCCCTCAGCTGCAGAAAGTGTTTGAA 1434

QY 1918 TCCATTGTGTAGGTTTAAACCCCTAGACTAAACGGTAATTTGTGTGAGAAAGAGTGAAC 1977
 DB 1435 GCCTTTGATA---ACTACCAACCCCAAGATGGTGGTGTGTTGTAGTTTCAAGAGAAAATCAGC 1491

QY 1978 ACCAGATTTTTCCTCAGTCTGGAGGAGAGCTTCAGAAATCCACTTCTCTGGAACAGTTATT 2037
 DB 1492 ACCAATCTGTACTTGTCTGCTGCTGATCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1551

QY 2038 GATGTAGAGGTTTACAGACAGCAATGGTATGATCTTTTATCGTGTAGCCAGGCTGTGAGA 2097
 DB 1552 GATCATACCAATAACAGCTGTGAGTGGTGGATTTCTACCTCTTTCGCCCATCATGTGCGA 1611

QY 2098 AGTGGTAGTGTCTTCTCCACACATTAATGATGCTCATATGACAAACAGCGGCTGGAAGCCA 2157
 DB 1612 CAGGGCTGTGGCATACACTACACTACATCTGTGTTCTGAAACTGCAAAATCTGAGCCCT 1671

QY 2158 GACCAATACAGGCTTACCTACAGCTGTGCCATCATTTTATACACTTGGCCAGGTGTC 2217
 DB 1672 GATCATATGAGAGGTGTGACTTTCAAACTATGCAATGCTGCTGCTGCTGCTGCTGCTGCT 1731

QY 2218 ATTTCGTGTTCTCTCTCTCTTCCAGTACGCGCCCAAGCTGGCTTTTCTTGTGGCCAGAGT 2277
 DB 1732 ATCCGAGTTCAGCTCTCTTGCAGAGTATGCCCAAGCTAGCTTCTCTGCTCGGACAGATT 1791

QY 2278 ATTCAAGAGAGCCAAATCTGTCTACTGTCAAAACCGCCCTTACTACTCTTAA 2328
DB 1792 TTGCATCATGAGCCAGCCATCCAGTGTGTGGGAACCTGTTCTTCTGTGTA 1842

RESULT 7

ID AB219511 standard; cDNA; 367 BP.
XX AC AB219511;

XX 23-JAN-2003 (first entry)

XX Group III cDNA cancer related clone SEQ ID NO:1937.

XX Human; cancer; tumour; therapy; diagnosis; CT antigen; CP antigen;
XX immune response; virology; immunology; microbiology; molecular biology;
XX recombinant DNA technology; gene; ss.

XX Homo sapiens.

XX WO200278516-A2.

XX 10-OCT-2002.

XX 28-MAR-2002; 2002WO-US10421.

XX 30-MAR-2001; 2001US-280255P.

XX 28-AUG-2001; 2001US-315563P.

XX 09-JAN-2002; 2002US-347313P.

XX (CORI-) CORIXA CORP.

XX Wang T, Wang S, Bangur CS, Gaiger A;

XX WPI; 2003-058387/05.

XX New immunogenic polynucleotides or polypeptides useful for diagnosing,
XX preventing and treating cancer expressing CT or CP mRNA antigens, and
XX in virology, immunology, microbiology, molecular biology and
XX recombinant DNA techniques

XX Claim 1; SEQ ID 1937; 207pp; English.

XX ABQ17575 to ABQ20506 represent isolated polynucleotide (I) sequences, and
XX ABP54446 to ABP54472 represent protein (II) sequences, from the present
XX invention. (I) and (II) have cytostatic activity and can be used in gene
XX therapy and vaccines. (I), (II), antibodies and compositions from the
XX present invention are useful for diagnosing, preventing and treating
XX cancer, which expresses CT or CP mRNA antigens. They are useful for
XX stimulating immune response. They can also be useful in virology,
XX immunology, microbiology, molecular biology and recombinant DNA
XX techniques.

XX N.B. The sequence data for this patent did not form part of the printed
XX specification, but was obtained in electronic format directly from WIPO
XX at ftp.wipo.int/pub/published_pct_sequences.

XX Sequence 367 BP; 118 A; 77 C; 91 G; 81 T; 0 other;

Query Match 15.7%; Score 365.4; DB 25; Length 367;

Best Local Similarity 99.7%; Pred. No. 1.6e-98;

Matches 366; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1291 AAAGTTACACAGCCATGGGCATCAATGAGAAAGCAATTAATGATTGAGTGATGAC 1350
DB 1 AAAGTTACACAGCCATGGGCATCAATGAGAAAGCAATTAATGATTGAGTGATGAC 60

QY 1351 AGAAGTGAAGCTTAAAGTCTTACAGCAAAAGTTCACAGCAGACACCCAGATAGTT 1410

DB 61 AGAAGTGAAGCTTAAAGTCTTACAGCAAAAGTTCACAGCAGACACCCAGATAGTT 120

QY 1411 GTCTGTCTGTGTCAGTAATCGGAAGACAAATACGATGCTATTAAAAATACCTGTGT 1470

DB 121 GTCTGTCTGTGTCAGTAATCGGAAGACAAATACGATGCTATTAAAAATACCTGTGT 180
QY 1471 ACAGATTGCCCTACCCCAAGTCAAGTGTGTGTGTCGCCCGAACCCTTAGCAAAACGAACT 1530
DB 181 ACAGATTGCCCTATCCCAAGTCAAGTGTGTGTGTCGCCCGAACCCTTAGCAAAACGAACT 240
QY 1531 GTCATGGCCATTGCTCAAAAGATTGCCCTACAGATGAATGCAAGATGGGAGGAGAGCTC 1590
DB 241 GTCATGGCCATTGCTCAAAAGATTGCCCTACAGATGAATGCAAGATGGGAGGAGAGCTC 300
QY 1591 TGAAGGTGGACATCCCCCTGAAGCTCGTGATGATCGTTGGCATCGATTCTTACCATGAC 1650
DB 301 TGAAGGTGGACATCCCCCTGAAGCTCGTGATGATCGTTGGCATCGATTCTTACCATGAC 360
QY 1651 ATGACAG 1657
DB 361 ATGACAG 367

RESULT 8

ABZ20005
ID ABZ20005 standard; cDNA; 367 BP.

XX AC ABZ20005;

XX 23-JAN-2003 (first entry)

XX Group III cDNA cancer related clone SEQ ID NO:2431.

XX Human; cancer; tumour; therapy; diagnosis; CT antigen; CP antigen;
XX immune response; virology; immunology; microbiology; molecular biology;
XX recombinant DNA technology; gene; ss.

XX Homo sapiens.

XX WO200278516-A2.

XX 10-OCT-2002.

XX 28-MAR-2002; 2002WO-US10421.

XX 30-MAR-2001; 2001US-280255P.

XX 28-AUG-2001; 2001US-315563P.

XX 09-JAN-2002; 2002US-347313P.

XX (CORI-) CORIXA CORP.

XX Wang T, Wang S, Bangur CS, Gaiger A;

XX WPI; 2003-058387/05.

XX New immunogenic polynucleotides or polypeptides useful for diagnosing,
XX preventing and treating cancer expressing CT or CP mRNA antigens, and
XX in virology, immunology, microbiology, molecular biology and
XX recombinant DNA techniques

XX Claim 1; SEQ ID 2431; 207pp; English.

XX ABQ17575 to ABQ20506 represent isolated polynucleotide (I) sequences, and
XX ABP54446 to ABP54472 represent protein (II) sequences, from the present
XX invention. (I) and (II) have cytostatic activity and can be used in gene
XX therapy and vaccines. (I), (II), antibodies and compositions from the
XX present invention are useful for diagnosing, preventing and treating
XX cancer, which expresses CT or CP mRNA antigens. They are useful for
XX stimulating immune response. They can also be useful in virology,
XX immunology, microbiology, molecular biology and recombinant DNA
XX techniques.

XX N.B. The sequence data for this patent did not form part of the printed
XX specification, but was obtained in electronic format directly from WIPO
XX at ftp.wipo.int/pub/published_pct_sequences.

XX Sequence 367 BP; 118 A; 78 C; 89 G; 79 T; 3 other;

Query Match 15.6%; Score 364; DB 25; Length 367;
 Best Local Similarity 99.2%; Pred. No. 4.1e-98;
 Matches 364; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1291 AAAGTTACACAGCCATGGCATGCAAAATGAGAAAGCAATAATGATTGAAGTGGATGAC 1350
 DB 1 AAAGTTACACAGCCATGGCATGCAAAATGAGAAAGCAATAATGATTGAAGTGGATGAC 60

QY 1351 AGAAGTGAAGCTACTTAAGAGTCTTACAGCAAAAGGTCACAGCAGACACCCAGATGTT 1410
 DB 61 AGAAGTGAAGCTACTTAAGAGTCTTACAGCAAAAGGTCACAGCAGACACCCAGATGTT 120

QY 1411 GTCGTCTGTCTCAAGTAATCGGAAGACAAATACGATGCTATTAAAAATACCTGTGT 1470
 DB 121 GTCGTCTGTCTCAAGTAATCGGAAGACAAATACGATGCTATTAAAAATACCTGTGT 180

QY 1471 ACAGATTCCTCCATACCCCAAGTCAGTGTGTGGCCGCAACTTAGGCAAAACAGCAAACT 1530
 DB 181 ACAGATNNCCCTACCCCAAGTCAGTGTGTGGCCGCAACTTAGGCAAAACAGCAAACT 240

QY 1531 GTCATGGCCATGCTACAAAGATTGCCCTACAGATGAACTGCAAGATGGGAGAGGCTC 1590
 DB 241 GTCATGGCCATGCTACAAAGATTGCCCTACANATGAACCTCAAGATGGGAGAGGCTC 300

QY 1591 TGGAGGTGGACATCCCTCGAGCTCGTGATGATCGTTGGCATCGATTGTACATGAC 1650
 DB 301 TGGAGGTGGACATCCCTCGAGCTCGTGATGATCGTTGGCATCGATTGTACATGAC 360

QY 1651 ATGACAG 1657
 DB 361 ATGACAG 367

RESULT 9
 AAH15960
 ID AAH15960 standard; cDNA; 2272 BP.
 XX
 AC AAH15960;
 XX
 DT 26-JUN-2001 (first entry)
 XX
 DE Human cDNA sequence SEQ ID NO:14558.
 XX
 KW Human; primer; detection; diagnosis; antisense therapy; gene therapy; ss.
 XX
 OS Homo sapiens.
 XX
 PN EPI074617-A2.
 XX
 PD 07-FEB-2001.
 XX
 PF 28-JUL-2000; 2000EP-0116126.
 XX
 PR 29-JUL-1999; 99JP-0248036.
 XX
 PR 27-AUG-1999; 99JP-0300253.
 XX
 PR 11-JAN-2000; 2000JP-0118776.
 XX
 PR 02-MAY-2000; 2000JP-0183767.
 XX
 PR 09-JUN-2000; 2000JP-0241899.
 XX
 PA (HELI-) HELIX RES INST.
 XX
 PI Ota T, Isogai T, Nishikawa T, Hayashi K, Saito K, Yamamoto J;
 XX
 PI Ishii S, Sugiyama T, Wakamatsu A, Nagai K, Otsuki T;
 XX
 DR WPI; 2001-318749/34.
 XX
 XX Primer sets for synthesizing polynucleotides, particularly the 5602
 PT full-length cDNAs defined in the specification, and for the detection
 PT and/or diagnosis of the abnormality of the proteins encoded by the
 PT full-length cDNAs -
 XX
 XX Claim 8; SEQ ID 14558; 2537pp + CD ROM; English.
 PS
 XX

CC The present invention describes primer sets for synthesizing 5602
 CC full-length cDNAs defined in the specification. Where a primer set
 CC comprises: (a) an oligo-dr primer and an oligonucleotide complementary
 CC to the complementary strand of a polynucleotide which comprises one of
 CC the 5602 nucleotide sequences defined in the specification, where the
 CC oligonucleotide comprises at least 15 nucleotides; or (b) a combination
 CC of an oligonucleotide comprising a sequence complementary to the
 CC complementary strand of a polynucleotide which comprises a 5'-end
 CC sequence and an oligonucleotide comprising a sequence complementary to a
 CC polynucleotide which comprises a 3'-end sequence, where the
 CC oligonucleotide comprises at least 15 nucleotides and the combination of
 CC the 5'-end sequence/3'-end sequence is selected from those defined in
 CC the specification. The primer sets can be used in antisense therapy and
 CC in gene therapy. The primers are useful for synthesizing polynucleotides,
 CC particularly full-length cDNAs. The primers are also useful for the
 CC detection and/or diagnosis of the abnormality of the proteins encoded by
 CC the full-length cDNAs. The primers allow obtaining of the full-length
 CC cDNAs easily without any specialised methods. AAH03166 to AAH13628 and
 CC AAH13633 to AAH18742 represent human cDNA sequences; AAH92446 to
 CC AAH95893 represent human amino acid sequences; and AAH13629 to AAH13632
 CC represent oligonucleotides, all of which are used in the exemplification
 CC of the present invention.

XX
 SQ Sequence 2272 BP; 609 A; 509 C; 585 G; 569 T; 0 other;

Query Match 14.2%; Score 330; DB 22; Length 2272;
 Best Local Similarity 51.6%; Pred. No. 1.9e-87;
 Matches 884; Conservative 0; Mismatches 810; Indels 18; Gaps 5;

QY 629 AAGAACAGTTTCCAAAGAACTAATAGTGTAGTGTCTTACCAAGTATAACAATAAGA 688
 DB 1 AGGATGAGTGTACTAAAGCTTCTGGTGGCAATATGTTATCATCCTGATATAACAATCGTA 60

QY 689 CATACAGAGTGGATGATATTGACTGGGACAGATCCCAAGAGCAGCCCTTTAAGAAAGCCG 748
 DB 61 CCTATCGTATTGAATGATGTGGATTGGAATAAGACTCCAAAGGATAGTTCACGATGCTG 120

QY 749 ACGGCTCTGAAGTCAGCTTCTTAGAATACTACAGGAAGCAATACAAACAGAGATCACCG 808
 DB 121 ATGGAAAGAGATCACATCTTGGATACCTACAGCAAAATTTATGGATCACAGTTAAGG 180

QY 809 ACTTGAAGCAGCTGTCTTGGTCAGCCAGCCCA---AGAGAAGCGGGGCGCTTGGGGGGA 865
 DB 181 AAGAGGACCAAGCCATTCCTGATTTCAGAGGCCAGTGCAGACAGAGGATAATCATGGGATGC 240

QY 866 CACTGCCAGGGCGCTGCATCTCATCTCAGTCTGCTATCTTACAGGCTCACTGATA 925
 DB 241 TGCTAAAGGGGAAATCTGCTGCTGCTGAGCTTTCTTTTATGACCCGGAATCCAGAGA 300

QY 926 AAATGCGTAATGATTTTAACTGATGAAAGACTTAGCCGTTTCATACAGACTAACTCCAG 985
 DB 301 AGATGAAGAGGACTTCAGAGCCATGAAGGATTTGGCTCAGCAATCAATCTGAGCCCA 360

QY 986 AGCAAGGCGAGCTGAAGTGGGAGCAGCTCATTCATTACATTCATAAAAGGATATGTTTC 1045
 DB 361 AGCAACACCATAGTGTCTTGGAAATGCTTGTGCAAAAGATTGCAAGAGACGAGGCCCA 420

QY 1046 AAGGGAGCTTCGAGACTGGGGTTTTCAGCTTTCATTCCTCACTTACTGCTCTTCAGGAA 1105
 DB 421 CCAATGAATGATGCTGCTGGGGCTCCGCTGCAAAAGGATGTACATAAGATTGAAGGAC 480

QY 1106 GAATTTTGAACAGAAAGATTCACCAAGGTGGAAAAACATTTGATTGATCAATCCCAAT 1165
 DB 481 GTGTTCTGCCAATGAAAGAAATTAACCTTA---AAAAATACTTCGTTTATCACATCTCAGG 537

QY 1166 TTCAGATTGGTCCAAAGAAACAGAGGTGCACCATTAATTAGTGTAGCCACTAGATA 1225
 DB 538 AACTAACTGGGTAAAGGAAGTAACCAAGAGACCCCTTCCATCTTGACTATCCCATGCAAT 597

QY 1226 ACTGGCTGTGTATCTATACGGAAGAAATTTATGACAGCCCAATTCATTGATACAAATC 1285
 DB 598 TCTGGGCACCTTTTATCCCAAGAGAGCAATGGACCAGGCTCGAGAACTGGTCAACATGT 657

1286	QY	TATTTAAAGTTACACCGCATGGCCATGCAAATAGAAAAAGCAATAATGATTGAAGT--	1344
658	Db	TGGAGAGATAGCGGCCCATTTGGCATGCGTATGAGCCACCGCGCTGGTTGAACTAA	717
1344	QY	-GGATGACAGAACTGAAGCGCTACTTAAAGAGTCTTTACAGCAAAAGGTCAACAGAG-	1396
718	Db	AGGATGACCGAATAGAGACTTATGTCCAGAACCATTTCAATCCAGTTAGGAGCTGAGGGGA	777
1397	QY	ACACCCAGATAGTGTCTGTCTGTGTCTCAAGTAATCGGAAGGACAAATACGATGCTATTAA	1456
778	Db	AGATACAGATGGTGTGTTTGCAATCATATGGGCCACACGTGATGATCTCTATGGGGCCATCA	837
1457	QY	AAAAATACCTGTGTACAGATTGGCCCTACCCCAAGTCAAGTGTGTGGTGGCCCGCAACCTTAG	1516
838	Db	AGAACTGTGCTGTGCAGTCCCGAGTGCCTCCAGGTTGTCAATGTTCGAACCATTTG	897
1517	QY	GCAAAACAGCAAACTGTGCATGGCCATTGCTACAAAGATTGCCCTACAGATGAATCTGAAGA	1576
898	Db	GTACAGCCCAACCGCTTCGGAGTGTGCCGTGCCCCAGAAAGATTTTACTTCAGATTAACTGTAAT	957
1577	QY	TGGAGGAGAGCTCTCGAGGCTGACATCCCCCTGAGAGCTCGTGAAGTATCGTTGGGATCG	1636
958	Db	TGGGTGTGAGCTCTGGGGAGTGATATCTCTGAAACAGTTAATGTGTGATCGGGATGG	1017
1637	QY	ATTGTTTACCATGACATCACAGCTGGGCGAGGTCAATCGCAGGATTTGTTGCCAGCATCA	1696
1018	Db	ATGTTTACCATGACCCAGTAGAGGATCGCTCCGTGGTTGCTTCGTGGCAGGATCA	1077
1697	QY	ATGAAGGATGACCCGCTGTTCTCAGCTGTCATTTTTCAGGATAGAGGACAGGAGCTGG	1756
1078	Db	ATCTCACCTTCACAAAATGATTTCCCGGTGGTGTTCAGATGCGCGCATCGAGAGATTG	1137
1757	QY	TGATGGGCTCAAAGTCTGCTCGCAAGCGCTCTGAGGGCTCGGAATAGCTGCAATGAGT	1816
1138	Db	TGACAGCCTTGAAGCTATGCTCTGTGGGCTCTTTAAAAAGTTTATAGGTGAACCACT	1197
1817	QY	ACATGCCCAACCCGATCATCTGTGTACCGCATGGCGTAGAGACGGCCAGCTCAAAAACAC	1876
1198	Db	GTCTACAGAGAAGATTGTGTGTACCGTATGGAGTGTCTGATGGCCAACTCGAAGACAG	1257
1877	QY	TGTTGAATACGAAGTCCACAGTTTGTGATTGTCTTAAATCCATTGTTAGAGGTTACA	1936
1258	Db	TTGCCAACTATGAGATTCTCTCAACTACAGAAGTGTTTTGAAGCTTTTGAGA---ATTATC	1314
1937	QY	ACCTAGACTAAAGGTAAATTGTGGTGAAGAAAAGAGTGAACACAGATTTTTTGTCTCAGT	1996
1315	Db	AGCCCAAGATGGTGGTGTGTGTAGTTCAGAAAGAAATCAGTACTAATCTATATCTGGCTG	1374
1997	QY	CTGGAGGAAGACTTCAGAAATCCACTCTCTCGAAACAGTTATTGTATGTAGAGGTTACAGAC	2056
1375	Db	CTCCTCAGAACTTTGTAACCTCCCACTCTCTGGAAGTGTGGTAGATCATACATTAACAGCT	1434
2057	QY	CAGAAATGGTAGACTTTTATTCGTGAGCCAGGCTGTGAGAAAGTGTGTTTCTCCCA	2116
1435	Db	GTGAGTGGGTGGATTTCTATCTTCTTGCCCATCATGTACGGCAGGGCTGTGGCATTTCTTA	1494
2117	QY	CACATTAACATGTCACTATGACAAACAGCGGCTGTGAAGCCAGACCATACACGCGTTGA	2176
1495	Db	CGCATATGTCTGTGTCTCAACACCGCAACCTGAGCCCTGTATGTATATGACAGAGCTGA	1554
2177	QY	CTTCAAGCTGTGCCACATCTATTACAACTGGCCAGGTGTCTTCTGTGTCTCTGTCTCTT	2236
1555	Db	CYTTCAAACTGTGGCACATGTACTGGAATTTGGCTGTGCCACCATCAGAGTTCCAGCTCCCT	1614
2237	QY	GCCAGTACGCCCAAGCTGGCTTTTCTGTTGGCCAGAGTATTTACAGAGAGCCAAATC	2296
1615	Db	GCAAGTATGCCCAAGCTAGCTTCTCTGTCAAGACACATCTTGATCATGTGACACGCA	1674
2297	QY	TGTCACTGTCAAAACCGCTTTTACTCTTAA	2328
1675	Db	TCCAGTGTCCGAGAACCTTCTTCTCTGTGA	1706

RESULT 10	
ABL10571	
ID	ABL10571 standard; cDNA; 2838 BP.
XX	
XX	ABL10571;
XX	
XX	26-MAR-2002 (first entry)
XX	
DE	Drosophila melanogaster expressed polynucleotide SEQ ID NO 26195.
XX	
KW	Drosophila; developmental biology; cell signalling; insecticide;
KW	pharmaceutical; gene; ss.
XX	
OS	Drosophila melanogaster.
XX	
PN	WO200171042-A2.
XX	
PD	27-SEP-2001.
XX	
PF	23-MAR-2001; 2001WO-US09231.
XX	
PR	23-MAR-2000; 2000US-191637P.
XX	
PR	11-JUL-2000; 2000US-0614150.
XX	
XX	(PEKE) PE CORP NY.
PA	
XX	
PI	Venter JC, Adams M, Li PWD, Myers EW;
XX	
DR	WPI; 2001-656960/75.
XX	
DR	P-PSDB; ABB66468.
XX	
PT	New isolated nucleic acid detection reagent for detecting 1000 or more
PT	genes from Drosophila and for elucidating cell signalling and cell-cell
PT	interactions -
XX	
PS	Claim 1; SEQ ID NO 26195; 21pp + Sequence Listing; English.
XX	
CC	The invention relates to an isolated nucleic acid detection reagent
CC	capable of detecting 1000 or more genes from Drosophila. The invention is
CC	useful in developmental biology and in elucidating cell signalling and
CC	cell-cell interactions in higher eukaryotes for the development of
CC	insecticides, therapeutics and pharmaceuticals. The invention
CC	discloses genomic DNA sequences (ABL16176-ABL30511), expressed DNA
CC	sequences (ABL01840-ABL16175) and the encoded proteins
CC	(ABB57737-ABB72072).
CC	The sequence data for this patent did not form part of the printed
CC	specification, but was obtained in electronic format directly from WIPO
CC	at ftp.wipo.int/pub/published_pct_sequences.
XX	
SQ	Sequence 2838 BP; 783 A; 659 C; 682 G; 714 T; 0 other;
	Query Match 13.3%; Score 310.6; DB 23; Length 2838;
	Best Local Similarity 49.1%; Pred. No. 1.4e-81;
	Matches 1134; Conservative 0; Mismatches 1129; Indels 46; Gaps 10
QY	52 AAACACAGTCTTTCAGGCATTATATAGTGTAAAGTTAAGCACTAACCAATTCGGCTGCACATCC 111
Db	431 AAGAAGAGAGTGGTTGGCACTCATATTACCGTCGAGCAAACTATTTTAAAGTATTAAAG 490
QY	112 CTTCCCCAGTGGCGCTTAATACGATATCAATTGACTATTAACCACTGATGGAAGCTCATGCTTTT 231
Db	491 CGTCCAAACTGGACCATCTACCAAGTACCGGTGCATTTTACGCTCATGTGGAGGCTACA 550
QY	172 AGACATCCGTTTCAGCTCTCTTTTTCACACGAGATCTAATTTGGAAGTGTCAATGCTTTT 231
Db	551 CGACATGCGACCGTCTTCTTGTATGAACATAAAGGGATCTCGGC--GGCTACATCTTTG 608
QY	232 GATGGAACGATATTATTTTACCTTAAAGACTACAGCAA-------AGGTTA 277
Db	609 ACGGAACCAATATGTTTGTGATCAATCAGTTCGAAGCTGTTCAAGATAGCCCTATGTTT 668
QY	278 CTGAAGTTTTTAGTACACCGGAATCGAGAGGATGTGAGGATPAAACGATCACTTTTAAACA 337

Db 669 TGGAACTGGTTACGAGAGCTGCTGGCGGAAACATTTGAAATTAGATCAAGCTGTG 728
Qy 338 ATGAACCTTCAACCTACATCAACCAACTGTGTGAGTTCTATATATATATTTTCAGGAGC 397
Db 729 GATCTGTGCAATCTACGGGATCGCGAGCAGTTTCAGGTCTCTAATCTCATACTGCGCAGG 788
Qy 398 TTTTGAATAATCATGAATTTGCAACAAATTTGGACGAATTTATTAACCCAAATGACCCAA 457
Db 789 CCAATGAGGGCTTAGACTTTGAAGCTGGTCTCGCGCTACTACGATCTCTCAAGCTAAGA 848
Qy 458 TTGATATTCAAAGTCACAGTTGGTGTGATTTGGCTCGCTTCACTACTTCCATCTTCAGT 517
Db 849 TTAATTTGGAGAATTTCCGATCGCAATTTAGCTTGGCTTACGACTTTCGATTCGCCAGC 908
Qy 518 ATGAAAACAGCATCTCTGCACTGAGCTTAGCCATAAGTCTTTCGAAATGAGACTG 577
Db 909 ACGAAAATGATATATTTACTTTTGTTCGAGATATGCCCAAGGTTATGCGAACTGAGACCT 968
Qy 578 TTTTGGATTTCAATGTTCAACTTTTATCATCAGACAGAAACATAAATTTCAAGACAG 637
Db 969 TGACATATTTTGTCCGATGCTATTTCCGAGAGTGAACGATTTATCAAGATACATTTAAGC 1028
Qy 638 TTTCCAAAGAACTAATAGTTAGTTTGTCTTACCAAGTATTAACATAAGACATACAGAG 697
Db 1029 GT-----GCAGTAATGGGTATGTTAATACCTAACGGATTTATTAACAAAACCTATCGCA 1082
Qy 698 TGGATGATATTGACTGGGACGAGATTTCCCAAGAGACCTTTTAAGAAAGCCGAGCTCTG 757
Db 1083 TTGAGATGTCGACTTTCAATCGAGCCATTTGTGCAAAATTT---AAAAACCAATGACGGTG 1139
Qy 758 AAGTCAGCTTTCTAGAAATCTACTACAGAGCAATTAACCAAGACCTTTTAAGAAAGCCGAGCTCTG 757
Db 1140 AGATTTCTGATGCTGATTAATAGAGCGATTAACCAATATTAATTAACAAAACCTATCGCA 1199
Qy 818 AGCTGTCTTGTGTCAGCCAGCCAGAGAGGCGGGCCCTGGGGGACACTGCCAGGCG 877
Db 1200 AACCTCTGCTCATGTCTC-GTCCGACAGACAAAATATTTGCTGTGGAATGACACGACC 1258
Qy 878 CTGCCATGCTCATCTGAGCTCTGTATCTTACAGGCTTAATGATTAATGATGCTGTAATG 937
Db 1259 ATA--ATGATCATTCGAGCTGGCAAGGCTTACGGAATGACGAGCTATGCGCGCTG 1316
Qy 938 ATTTTAACGTGATGAAGACTTAGCCGTTTATACAGAGCTAACTACTCCAGAGCAAGGACG 997
Db 1317 ACTTTAGAGCTTGAAGCCATGAGTGAACATACAGGCTGAATCCAGATGCTGCGCATCG 1376
Qy 998 GTGAAGTGGGACGACTCATTTGATTAATTTCAATAAAGATATGTTCAAGGGAGCTTC 1057
Db 1377 AACGCTTCCGATGTTCAACAGCGCTTAAAGTCTGTGTAACAGAGTGTAGAGACCTTA 1436
Qy 1058 GAGACTGGGTTTGAAGCTTTGATTTCAACTTACTGCTCTTCTCAGAGAAATTTTGCAAA 1117
Db 1437 AGTCTGGAAACATCCAGCTGGAAGCTTGGTGGAGATTCACAGCAGCTGTTGGCAC 1496
Qy 1118 CAGAAAAGATTCACCAAGGTGGAAAAACATTTGATTAACATCCCAATTTTGAGATTCGT 1177
Db 1497 CGGAAAATATTTTCGGCAACCAAAATATTTGTAAGCAGCTCGCGCGGATTTGA 1556
Qy 1178 CCAAGAAACAGAGGTGCAATTAATTTAGTTTGAAGCCACTAGATTAAGCTGTTGTA 1237
Db 1557 CCAATGAGTTTCCGACCTGTTTCGATGTTTAAAGCTGCAATATCAATAGGTGTTATGTA 1616
Qy 1238 TCTATACCGAAGAAATATGAGCAGAGCAATTCATGATCAAAATCTATTTAAGTTA 1297
Db 1617 TCACACCGAGTGAATCTGCGCGAGACCCAGGAATTTGTGAGATGTCATCCGACAG 1676
Qy 1298 CACAGCATGGGCTGCAATGAGAAAGCAATTAATGATTTGAAGT---GGATGACGAA 1354
Db 1677 CCAGCAGCATGAAGTGAACCTCTGCAATCAATATATAGAGAAATTCGGATGACCGCA 1736
Qy 1355 CTGAAGCTTACTTAAGTCTTTACAGCAAAAGGTACAGCAGACACCCAGATAGTTGCT 1414
Db 1737 ACGGCACTTACTCCCAAGCGATGCAACAGCGGAGCTTAACGATCCAGATGCTGATG 1796

Qy 1415 GTCTGTTGTCAAGTAAATCGGAGGACAAATACGATGCTATTAAAAATACCTGTGTACAG 1474
Db 1797 TTGTCAATGAGATCTCCAAATGAAGAGAATATAGCTGCATTTAAAAACGACGTCGGTG 1856
Qy 1475 ATTGSCCTTACCCCAAGTCAAGTGTGT-----GGTGGCCCGAACCTTAGGCAAAACAG 1524
Db 1857 ACAGACCGGTGCCATCGCAGGTGTGACATAAAAGTTATCGCGCTCGACACGCAAAAAC 1916
Qy 1525 CAAACGTG--TCATGSCCAATTCCTACAAAGATTGCCCTACAGATGAATTCAGATGGGAG 1582
Db 1917 CAACTGGGCTGATGTCGATCGCCACAAAGGTGTTATTCAANTGAACGCCAAATGATGG 1976
Qy 1583 GAGAGCTCTGAGAGGTGACATCCCTGAAAGCTGCTGATGATCTTTGGCATTCGATGTT 1642
Db 1977 GAGCTCCCTGSCAGGTAGTGTATCCCTCCAGGTCIGATGACTTGTGTTTCGATGCT 2036
Qy 1643 ACCATGACATGACAGCTGGGCGAGGTCAATCGCAGGATTTGTTGCCAGCATCAATGAAG 1702
Db 2037 GCCATTCCCGAAGAAATAAGAACCAAGTCATATGGGGCTTTGTTGCCAACCATGGACCAGA 2096
Qy 1703 GGATGAC---CGGCTGTTCTCACGCTGCATATTTCAGATAGAGGACAGAGCTGGTAG 1759
Db 2097 AGGAGTCTTTCCGCTACTTTCTCCACCGTAAACGACACATTAAGGGCCCAAGAGTTGTCG 2156
Qy 1760 ATGGCTCAAAAGTCTGCTGCAAGCGGCTCTGAGGGCTTGGAAATAGCTGCAATGAGTACA 1819
Db 2157 AGCAGATGTGGTGAACATGCGTGGCGCTGAGGTGCTGATCAGGAGCAACACCGTTCTT 2216
Qy 1820 TGCCAGCGGATCATGTTGTTACCGGATGCGGTAGGAGACGGCCAGCTGAAAACACTGG 1879
Db 2217 TGCCAGAGCGCATTTCTTTCTTTCGCGACGCTGTTGGGATGCTGAGCTTACAGAGTGG 2276
Qy 1880 TGAATCAAGAGTGCACAGTTTTTGGATTTCTTAAATCCATTTGGTAGAGTTTACAACC 1939
Db 2277 TAAACAGGAGGTGAACACCCCTAAAGGACAGGCTCGAGAAATTTACAATCAGCTGGCA 2336
Qy 1940 CTAGACTAAACGTTAATTTGTTGTAAGAAAGAGTGAACACAGAGTTTTTTGCTCAGTCTG 1999
Db 2337 AACAGGAGGCTGTCGCATGACATTTATTTATTTGATCCAGCGCATTAATTTCTGCTACT 2396
Qy 2000 GAGGAAGCTTCAGATTCCTTCTGGAACAGTTATTTGATGTAGAGTTTACCAAGACAG 2059
Db 2397 TTACTGGGCTATCGCAACCCAGTTCGCGGCACTGTAGTCGATGACGTTATTACCTTGCAG 2456
Qy 2060 AATGCTATGACTTTTATCGTAGCCAGGCTGTGAGAGTGTGAGTGTGTTCTCCACAC 2119
Db 2457 AGCCTAGACTTCTTCTAGTGTCCAGGCTGTTGCGATAGGAACCTGTGCGCTTACCA 2516
Qy 2120 ATTCAATGTCACTATGACAAACAGCGGCTTGAAGCCAGACCAATACAGCGCTTGACCT 2179
Db 2517 GCTACAATGTTATTTCTGACAACTGGGCTTAAACGCGGATAAGCTGAGATGCTCTCT 2576
Qy 2180 ACAGCTGTGCCATCTATTAACCTGGCCAGGTGTCTATGTTCTGCTGCTCTGCTGCC 2239
Db 2577 ATAAGATGACCCATGATGTACTACAATTAACAGCGGAACCATACGAGTCCCGCTCTGCC 2636
Qy 2240 AGTACGCCCAACAGCTGGCTTTCTTCTGTCGACAGATTTTACAGAGAGCCAAATCTGT 2299
Db 2637 ATTAAGCCCAACAAATTTGCTTCTGTCGAGCGGAATCCATTAATCGTGGCTTACAGAG 2696
Qy 2300 CACTGTCAAAACCGCTTTTACTACTCTAA 2328
Db 2697 GACTGCAAGATCAATTTGTTTGTAA 2725

RESULT 11

AAZ13404

ID AAZ13404 standard; cdna; 300 BP.

XX AC

XX AAZ13404;

DT 12-OCT-1999 (first entry)

XX DE Human gene expression product cDNA sequence SEQ ID NO:873.
XX DE Human; gene; gene expression product; diagnosis; therapy; probe;
XX DE detection; mapping; tissue typing; profiling; forensic; cancer;
XX DE genetic analysis; colorectal cancer; breast cancer; lung cancer; ss.
XX OS Homo sapiens.
XX PN WO9938972-A2.
XX PD 05-AUG-1999.
XX PF 28-JAN-1999; 99WO-US01619.
XX PR 03-APR-1998; 98US-0080666.
XX PR 28-JAN-1998; 98US-0072910.
XX PR 24-FEB-1998; 98US-0075954.
XX PR 31-MAR-1998; 98US-0080114.
XX PR 03-APR-1998; 98US-0080515.
XX PA (CHIR) CHIRON CORP.
XX PA (HYSE-) HYSEQ INC.
XX PI Crkvenjakov R, Dickson M, Drmanac R, Drmanac S;
XX PI Escobedo J, Garcia PD, Garcia V, Giese K, Innis MA;
XX PI Jones WL, Kassam A, Kennedy GC, Kita D, Labat I;
XX PI Lamson G, Leshkowitz D, Pot D, Randazzo F, Reinhard C;
XX PI Stache-Crain B, Sudduth-Klinger J, Williams LT;
XX DR WPI; 1999-494092/41.
XX XX Novel human genes and their expression products which are
XX PT differentially expressed in different cell types
XX PS Claim 1; Page 862-863; 2479pp; English.
XX XX The present invention describes a library of human polynucleotides
XX CC comprising the sequences given in AA212532 to AA21779, also described is
XX CC a method of detecting differentially expressed genes correlated with the
XX CC cancerous state of a mammalian cell, comprising detecting at least one
XX CC differentially expressed gene product in a test sample from a cell
XX CC suspected of being cancerous, where the gene product is encoded by one
XX CC of the 5248 polynucleotide sequences given in AA212532 to AA21779. The
XX CC polynucleotides can be used as a source of primers and probes, which can
XX CC be used for a variety of purpose, e.g. detection of expression levels,
XX CC mapping, tissue typing or profiling, forensics, genetic analysis and
XX CC detection of polymorphisms. Polypeptides encoded by the polynucleotides
XX CC can be used for raising antibodies for experimental, diagnostic and
XX CC therapeutic purposes. The polynucleotides may also be used to construct
XX CC arrays for diagnostics (which may be used to determine function of an
XX CC encoded protein); and to detect differences in expression levels between
XX CC two cells (e.g. to identify abnormal or diseased tissue in a human, to
XX CC identify a genetic predisposition or susceptibility to a disease such as
XX CC cancer). The polynucleotides of the invention are especially used in the
XX CC diagnosis, prognosis and management of colorectal cancer, breast cancer,
XX CC and lung cancer. The polynucleotides can also be used to screen for
XX CC peptide analogues and antagonists.
XX SQ Sequence 300 BP; 75 A; 70 C; 91 G; 64 T; 0 other;
Query Match 12.3%; Score 286.4; DB 20; Length 300;
Best Local Similarity 99.3%; Pred. No. 6e-75;
Matches 298; Conservative 0; Mismatches 1; Indels 1; Gaps 1;
QY 1485 CCCAAGTCAGTGTGTGGTGGCCGACCTTAGCAACAGCAAGCTGTATGCGCATTCG 1544
Db 1 CCCAAGTCAGTGTGTGGTGGCCGACCTTAGCAACAGCAAGCTGTATGCGCATTCG 60
QY 1545 TACAAAGATTGCCCTACAGATGAACCTAGCAATGGGAGAGAGCTCTGGAGGGTGGACAT 1604
Db 61 TACAAAGATTGCCCTACAGATGAACCTAGCAATGGGAGAGAGCTCTGGAGGGTGGACAT 120

QY 1605 CCCCCTGAAGCTCGTCGATGATCGTTGGCATCGATTGTTACCATGACATCAGCTGGGGC 1664
Db 121 CCCCCTGAAGCTCGTCGATGATCGTTGGCATCGATTGTTACCATGACATCAGCTGGGGC 180
QY 1665 GAGGTCAATCGCAGGATTGTTGCCAGCATCAATGAAGGATGACCCCGCTGTTCTCAGC 1724
Db 181 GAGGTCAATCGCAGGATTGTTGCCAGCATCAATGAAGGATGACCCCGCTGTTCTCAGC 240
QY 1725 CTGCATATTTCAGGATAGAGGACAGAGCTGTGTAGTAGGGCTCAAAGTCTGCTGCAAGC 1784
Db 241 CTGCATATTTCAGGATAGAGGACAGAGCTGTGTAGTAGGGCTCAGAG-CTGCTGCAAGC 299

RESULT 12
ABL10535
ID ABL10535 standard; cDNA; 3524 BP.
XX AC ABL10535;
XX DT 26-MAR-2002 (first entry)
XX XX Drosophila melanogaster expressed polynucleotide SEQ ID NO 25087.
XX DE Drosophila; developmental biology; cell signalling; insecticide;
XX KW pharmaceutical; gene; ss.
XX OS Drosophila melanogaster.
XX XX WO200171042-A2.
XX PN 27-SEP-2001.
XX PD 23-MAR-2001; 2001WO-US09231.
XX PF 23-MAR-2000; 2000US-191637P.
XX PR 11-JUL-2000; 2000US-0614150.
XX XX (PEKE) PE CORP NY.
XX PI Venter JC, Adams M, Li PWD, Myers EW;
XX DR WPI; 2001-656860/75.
XX DR P-PSDB; ABB66432.
XX XX New isolated nucleic acid detection reagent for detecting 1000 or more
XX PT genes from Drosophila and for elucidating cell signalling and cell-cell
XX PT interactions -
XX PS Claim 1; SEQ ID NO 26087; 21pp + Sequence Listing; English.
XX CC The invention relates to an isolated nucleic acid detection reagent
XX CC capable of detecting 1000 or more genes from Drosophila. The invention is
XX CC useful in developmental biology and in elucidating cell signalling and
XX CC cell-cell interactions in higher eukaryotes for the development of
XX CC insecticides, therapeutics and pharmaceutical drugs. The invention
XX CC discloses genomic DNA sequences (ABU16176-ABU30511), expressed DNA
XX CC sequences (ABU01840-ABU16175) and the encoded proteins
XX CC (ABB57737-ABB72072).
XX CC The sequence data for this patent did not form part of the printed
XX CC specification, but was obtained in electronic format directly from WIPO
XX CC at ftp.wipo.int/pub/published_pct_sequences.
XX SQ Sequence 3524 BP; 1053 A; 741 C; 761 G; 969 T; 0 other;
Query Match 11.3%; Score 263.4; DB 23; Length 3524;
Best Local Similarity 47.4%; Pred. No. 2.1e-67;
Matches 1088; Conservative 0; Mismatches 1176; Indels 29; Gaps 9;
QY 46 GAATCAAAAAACAGTTCTTCAGGCAATTATAGTAAAGTTAAGCACTAACCAATTTCCGGCTG 105
Db 342 GTATCCCAAGAGGAGGACCGATGGCGTCCCGGTTCATGTCGACGACGAACATTTTCCGATTA 401
QY 106 ACATCCCGTCCCGAGTGGGCTTATATACATATACATTAACCACTGATCGAA 165

Db 402 AAAACCAAGCCGGAATGGCGGATCGTTTCAATTAACGTTGGAGTGTGAGCCGAGCATCGAG 461
Qy 166 GCCAGAAGACTCCGTTCACTCTTCTTTTCAACGAGAGATCTAATTCGAAAGTGTGAT 225
Db 462 AATCCTCGTGTCCGATGGAGTTTGTCCATCATCTGCTAACTTCTGGATCAGGCTAT 521
Qy 226 GCTTTTGATGGAACGATATATTTTCTAAAGACTACAGAAAGGTTTACTGAAGTT 285
Db 522 CTATTCGATGGAGTCAACTGTTTCCACCACAGGAAATTCGAGCAGGAATCACTGGTGTCT 581
Qy 286 TTTAGTAAGACCCGGAATGGAGAGATGTGAGATAACGATCACTTTTAAACAAATGAATCT 345
Db 582 AGCGAAGTTCGAAGCTGACATTTGAATTAACAGATATCCATTAAGTTCTGTTGATTCATA 641
Qy 346 CCACCTCATATCCAACTTTGTTGAGTTTCTTAATATATTTTTCAGGAGGCTTTTGAAA 405
Db 642 TCGTGTGCTAGCCCCGCTTTTTCGAAGTCTTAAATCTAATATTTGCGCGCTCCATGAAG 701
Qy 406 ATCATGAATTTGCAACAAATTCGACGAATATTAATTAACCCAAATGACCCAAATGATAT 465
Db 702 GGCCTAAATTTGGAATTAAGTTGGCCGTAACTCTTTTGATCCCGAGCTAAAGATCGAAATA 761
Qy 466 CCAAGTCAACAGTTGGTGAATTTGGCTGGCTTCACTTCTTCACTTCCATCTTCAAGTATGAAC 525
Db 762 AGGAGTTCAAAATGAGACTATGGCGGCTATGACATCGAATTCGTCAGACGCAAAA 821
Qy 526 AGCATATGCTCTGCACTGACGTTAGCCATAAAGTCTTTCGAAGTGAAGTCTTTTGGAT 585
Db 822 GATATTTTATGGGCACCGAAATACTCACAAAGTTATGCGCACCGAGAGATCTACGAC 881
Qy 586 TTCATGTTCACTTTTATCATCAGACAGAACTAATAATTTCAAGAACAGTTTCCAAA 645
Db 882 AATAAGCAGGTTGCTCACAAATCCGGCTCGTCTATCAGGAGAAAGTACGGGT-----A 935
Qy 646 GAACTAATAGGTTAGTTGTTCTTACCAGTATAAATAAAGACATACAGAGTGGATGAT 705
Db 936 AATGTTTGGACTGATGTTCTTACGATTAACAATAACAGAACTTATCGTATCAATGAT 995
Qy 706 ATTGACTGGGACCAAAATCCCAAGAGACCTTTTAAAGAAAGCGAGGCTCTGAAGTCAAG 765
Db 996 GTGACTTTGGACAACTCCGAAATCAACATT--CAGTTGCAAGGTTAGAGATATCAGT 1052
Qy 766 TTCATTAGATATACAGAGCAATTAACACCAAGAGATCACCGACTTGAAGCAGCCTGTC 825
Db 1053 TTCTGGGAATATCTCTCACTAAATATAATATACGATTCGCGACCAATCAATCAGCCGCTG 1112
Qy 826 TTGTTCAAGCCCAAGAGAGGCGGCGCTTGGGGGACACTCCAGGGGCTGCCCATG 885
Db 1113 CTGATCTCCAAA--AATAGGACAAGGCTCTAATAAATCAAGCTAGCGAATTAGTGTA 1169
Qy 886 CTCATTCCTGAGCTCTGTTATCTTACAGTCTTAAGTATAAATTCGTAATGATTTTAAAC 945
Db 1170 CTAATTCCTGAGCTCTGCGAGTGTACTGGGCTCAATGCGGAGATGCGCTCAAACTTTTCA 1229
Qy 946 GTGATGAAGACTTAGCGTTTCATACAGACTAATCTCCAGAGCAAGCAGCGTCAAGTG 1005
Db 1230 CTTATGCGTGCCATGAGCAGTTATCGGGAATGAACCCCAACACGCACTGATCGATG 1289
Qy 1006 GGAGACTCATTTGATTAATTCATATAAAGCATATGTTTCAAGGAGGACTTTCGAGACTGG 1065
Db 1290 CGCGCTTTTAAACCCGTTTACAAACACTCCAGAAAGTGTGAGGTTCTTGGAGACTGG 1349
Qy 1066 GGTTTGACTTTGATTCGAATCTGCTTCTCAGGAAGAAATTTGCAACAGAAAG 1125
Db 1350 AACATGGAACGTGCAAGAAAGTACAGAAAGTACAAAGCCGGAATAATTTGGACAGCAGAAC 1409
Qy 1126 AT---TCACCAAGTGGAAACATTTGATTAACATCCAAATTTGAGATTTGTTGCTCAA 1181
Db 1410 ATCGTGTTCATATGAAAGGTTCCCTGCTGGAGAAACCGTGAATTTGGCAAGGCACTTC 1469
Qy 1182 AGAAACAGAGGTGCACCAATTAATTTAGTTGTTAAGCCACTAGATAACTGGCTGTTGATCTA 1241

Db 1470 AGAGACCAAGAGATGCTTTTACCCTCCGAGCGATGGC-CTCGATCGTTGGGCTGTCTATCGC 1528
Qy 1242 TACCGGAAGAAATATGAAGCAGCAATTTATTTGATACAAATCTATTATTAAGTTTACACC 1301
Db 1529 GCCGCAAGGAATTTCCCATGAACCTCCGAACCTTACTTGACTCTTTGTATAGCAGCTAG 1588
Qy 1302 AGCCATGGCCATGCAATGAGAAAAGC---AATATGATTTGAAGTGGATGACAGAACTGA 1358
Db 1589 TGAATGGGTCTTAGAATTCGAAGCCCGCAAGATTCATAATTTTATGATGATGCGACTGG 1648
Qy 1359 AGCTTACTTAAGTCTTACAGCAAAAGGTTCAGAGCAGACCCAGATGTTGCTGCT 1418
Db 1649 AACTTATGTGAGAGCAATTTGTCGCTCAGATCCCAAACTTATATTATGCT 1708
Qy 1419 GTTGTCAAGTAATCCGAAGGACAAATACGATGCTTATTAATAAATACCTGTCTACAGATTG 1478
Db 1709 CGTACCCNATGATAACGCCGAAGATTAATCATCAATCAAAAAGAGAGATGATGTTGACAG 1768
Qy 1479 CCCTACCCAAAGTCAAGTGTGTGGCCGCAACCTTAGGCAAAACAGCAAACTGTCAAGC 1538
Db 1769 GCGGTGCGCAACTCA---AGTTGTGACCTTAAACCGAACCAAGAACCGTAGCTTATGAG 1825
Qy 1539 CATTGCTACAAAGATTTGCCCTACAGATCAACTGCAAGATGCGAGAGAGCTCTGAGAGGT 1598
Db 1826 CATTGCCACCAAAATAGCAATCCAACTGAATTCGAAGTTGGGATATACACCTCGATGAT 1885
Qy 1599 GGACATCCCCCTGAAGCTCGTGATGATCGTTGGCATCGATTTGTTACCATGATGACAGC 1658
Db 1886 CGAATACCTTGTCCGACTGATGACAAATTTGGCTTTGACATTCGGAAGACACACGAGA 1945
Qy 1659 TGGCGGAGGTCAATTCGAGAGATTTGTTGCAAGATCAATGAAGGATGACCCGCTGCT 1718
Db 1946 TCGAAGAGGGCTTACGAGCAATTTGCTCAATGATCTACAGCAAACTCCACGTA 2005
Qy 1719 CTCACGTGCTATTTTCAAGATGAGGACAGGAGCTGTAGATGGCTCAAGTCTGCT 1778
Db 2006 CTTCAAGACAGTACGAGGTGACGCGCTTGTGATGCTCGCTAACACCTTTTGGCGAT 2065
Qy 1779 GCAAGC---GGCTCTGAGGGCTTGGATGATGCTCAATGAGTACATGCCCGCCGATCAT 1835
Db 2066 GATAGCAAGGCTCTGCGCAATATCAATGATGATGAGAGCTGCCATCTCGAATGCT 2125
Qy 1836 CGTGTACCGGATGGCTGAGAGACGCGCACTGAAACACACTGTGTGAACACTGCAAGTGC 1895
Db 2126 ATTTTATCGAGACGCTGTGAGCTCCGCTCTCTAAAGCAGCTTTTGAATTTGAAGTCAA 2185
Qy 1896 ACAGTTTTTGAATGTTCTAAATCCATTGTTAGAGTTTCAACCTAGATTAACGGTAA 1955
Db 2186 GGACATCATTTGAGAGTTGAAACTGAATAGCCCGGCTCCAGCTAAGCCCGCAAT 2245
Qy 1956 TGTGTTGAAGAAAGAGTGAACACAGATTTTGTCTCAGTCTGGAGGAGACTTCAGAA 2015
Db 2246 AGCT---TATATTGTGTAACCATCCAGATCCATGAACACGCGCTTCTTCTCAACGACAAA 2302
Qy 2016 TCCACTTCTCGAACACTTATGATGTAGAGTTTACAGACCAAGATGATGATGCTTTT 2075
Db 2303 TCCCTCCCTGGTACTATGATGATGATGATGATGATGATGATGATGATGATGATGAT 2362
Qy 2076 TATCGTCAAGCAGGCTGTGAGAGTGTGATGTTTCTCCACACATTAACATGATGATCTA 2135
Db 2363 TCTGTTCTCGCAACAGTTCTGTAGAGTACAGTGTGCGGACGAGTACATGATGTTCTTA 2422
Qy 2136 TGAACACAGCGGCTGAAGCCAGACCAATACAGCGCTTGACCTCAAGCTGTGCCACAT 2195
Db 2423 TAGCAGCATGGTCTCTCCGCGAGAAATGCAAAACTTACGTACAGATGTGCCACTT 2482
Qy 2196 CTATTACAACTGGCAGGTGTCTTCTGTTCTGCTTCTGCTTCTGCTGCTGCTGCTGCT 2255
Db 2483 GTACTACAAATTTGTCGGGCAACCAACAGAGTCCAGAGTGTGCGAGTACCGTAAGAGCT 2542
Qy 2256 GGCCTTTCTGTTGGCCAGAGTATTCAGAGAGGCAAAATCTGCTCACTGTCAACCGCT 2315
Db 2543 AGCTACCTCTGTTGGTACGAATTTGCACTCTATTCGCAAAACGCGCTCGAAAGAGATT 2602

```

QY 2316 TTACTACCTCTAA 2328
DB 2603 TTATATCTATAA 2615

RESULT 13
ID AAA07586
XX AAA07586 standard; DNA; 3047 BP.
AC AAA07586;
XX 29-AUG-2000 (first entry)
DT XX
DE Drosophila piwi gene.
XX
KW Piwi family protein; piwi; miwi; hiwi; gene therapy; tissue dystrophy;
KW anaemia; immunodeficiency; male infertility; Drosophila; ds.
XX
OS Drosophila sp.
XX
FH Key Location/Qualifiers
FT CDS 84..2615
FT /product= piwi
FT /transl_except= (pos:120..122; aa:Xaa)
FT /transl_except= (pos:399..401; aa:Xaa)
FT /transl_except= (pos:2436..2438; aa:Xaa)
FT /note= "Xaa= Leu or Ile"
XX
PN WO200032039-A1.
XX
XX 08-JUN-2000.
XX
XX 03-DEC-1999; 99WO-US28764.
XX
XX 04-DEC-1998; 98US-0110901.
XX
XX (UYDU-) UNIV DUKE.
XX
XX Lin H;
XX
XX WPI; 2000-412085/35.
XX P-PSDB; AAY90233.
XX
XX Piwi family nucleic acids, polypeptides, and antibodies, useful in gene
XX therapy of diseases such as cancer and in various research and
XX diagnostic applications -
XX
XX Claim 19; Page 171-175; 201pp; English.
XX
XX This sequence encodes the Drosophila piwi family protein, designated
XX piwi. The piwi family nucleic acids and polypeptides are used in gene
XX therapy of diseases such as cancer and also in various research and
XX diagnostic applications. The sequences can also be used to treat
XX tissue dystrophy, anaemia, immunodeficiency, and male infertility.
XX
XX Sequence 3047 BP; 936 A; 664 C; 672 G; 772 T; 3 other;

Query Match 11.0%; Score 256.6; DB 21; Length 3047;
Best Local Similarity 47.2%; Pred. No. 2.1e-65;
Matches 1083; Conservative 0; Mismatches 1181; Indels 29; Gaps 9;

QY 46 GAATCAAAAACAGTTCTTCAGGCATATATAGTAGTTAAGCTATACCATTCACCATTCGCGCTG 105
DB 342 GTATCCAAAGAGGAAACCGATGGCGTCCCGGTCAATGCTGAGAGCAACTTTTTTCGANTA 401
QY 106 ACATCCCGTCCCGAGTGGGCTTATATACAGTATACATTCAGTATACCATTCAGTATACCATTCAGT 165
DB 402 AAAACCAAGCGGAATGGCGATGTTTCATATACAGTGGAGTTTGGCGACCATCGAG 461
QY 166 GCCAGAAAGATCCCGTTCAGTCTCTTTTCAACACGAAAGATCTAAATGGAAAGTGTCAAT 225

```

```

DB 462 AATCCTCGTGTCCGTATGGGAGTTTGTCCAAATCATGCTAACCTTCTGGGATCAGGCTAT 521
QY 226 GCTTTTGATGGAACGATATATTTTACCTAAAGACTACAGCAAAAGGTTTACTGAAGTT 285
DB 522 CTATTCGACGGAAGTCAACTGTTTCCACCACAGGAAATTCGAGCAGGAAATCACGGTGTCT 581
QY 286 TTTAGTAAGACCCCGAATGGAGAGGATGTGAGATAGCATCTTTTAAACAAATGAACATT 345
DB 582 AGCGAAAGTTCGAAGCTGGACATTAAGATATCCATAAAGTTTCGTTGGATTCATA 641
QY 346 CCACCTACATCACCAACTTGTGTTCAGTTCTATAAATATATTTTTCAGGAGCTTTTGA 405
DB 642 TCGTGTCTGAGCCCGCTTTTTCAGTCTTAATCTAATATTCGCGCTCCATGAAG 701
QY 406 ATCATGAATTTGCAACAAATTTGCAAGAAATTTATTAACCCAAATGACCCAAATGATTT 465
DB 702 GGCTTAATTTGGAATTTAGTTGGCCGTAATCTCTTTGATCCCGAGCTAAGATCGAATA 761
QY 466 CCAAGTCACAGGTTGGTGAATTTGSCCTGGCTTCACTACTTCCATCTCTCATGATGAAC 525
DB 762 AGGAGTTCAAAATGGAGCTATGGCCGGGTATGAGACATCGATTCGTCAGCAGCAAAA 821
QY 526 AGCATCATGCTCTGCACTGAGCTTAGCCATAAAGTCTCTGAAAGTGAGACTGTTTGGAT 585
DB 822 GATATTTTATGGGACCCGAAATAACTCAAAAGTTATGCGCACCGAGACGATCTACGAC 881
QY 586 TTCAATGTTCAACTTTTATCATCAGACAGAAAGACATTAATTTCAAGAAACAGTTTCCAA 645
DB 882 ATAATGCGAGTTTGTCTCACAAATCGGGCTCGTCATCAGGACGAAAGTACGGGT -A 935
QY 646 GAACTAATAGTTTATGTTGTTTACCAAGTATAACAATAAGACATACAGAGTGATGAT 705
DB 936 AATGTTTTCAGCTTGATGTCTTCGATTAAGATTAAGACATTAATTCGATCAATGAT 995
QY 706 ATTGACTGGGACCGAATCCCAAGAGACACCTTTAAGAAAGCCGACGGCTCTGAGTCA 765
DB 996 GTCGACTTTGGACAAACTCCGAATCAACATT--CAGTTGCAAGGCTAGAGATATCAGT 1052
QY 766 TTCTTAGAATACTACAGGAGAGAAATACACCAAGAGATACCGACTTGAAGACGCTGTCT 825
DB 1053 TTCGTGGAATACTATCTCACTAAATATAATATACGCAATTCGCGACCACAATACGCGCTG 1112
QY 826 TTGTTGAGCCAGCCCAAGAGAGGCGGGGCGCTGGGGGACACTGCCAGGCGCTGCCATG 885
DB 1113 CTCATTTCCAAA--AATAGGACAGGCTCTAATAAATACCGCTAGCGAATTAAGTGGTA 1169
QY 886 CTCATTTCTGAGCTCTGCTATCTTACAGGTCTAACTGATAAAATGGTATGATTTTAA 945
DB 1170 CTAATTTCTGAGCTCTGCGAGTGACTGGGCTCAATGCCGAGATCGCTCAAACTTTCAG 1229
QY 946 GTGATGAAGACTTAGCCGTTTATACAGACTTAATCTCCAGAGCAAAAGGCGAGCTGAAGTG 1005
DB 1230 CTTATCGTGCCTACGACAGCTTATACGGAATGAACCCCAACACGCACTGATCGATTG 1289
QY 1006 GGACGACTCATTCATTACATTAATAAAGCAATGTTCAAGAGGAGCTTCGAGCTGG 1065
DB 1290 CGGCTTTTACCAACCGTTTACAAAACATCTCCAGAAAGTGTGAAGGCTTTGAGAGACTGG 1349
QY 1066 GGTTCGAGCTTTGATTTCCAACTTACTGTCTCTCAGGAAGAATTTTCAAAACAGAAAAG 1125
DB 1350 AACATGGAATCGACAAGAACGTCACAGAAGTACAGGCGGATTAATTGACAGCAGAAC 1409
QY 1126 AT----TCACCAAGGTGGAAAAACATTTGATTAATTCACATTCACATTTGCGAGATTG 1181
DB 1410 ATCGTGTTCATAATGAAAAGGTTCTGCTGGGAAAAAGCTGATTTGGCAAGGCACTTC 1469
QY 1182 AGAAACAGAGGTGCACCAATTAATTAAGTGTTAAGCCACTAGATAACTGGCTGTGTCTA 1241
DB 1470 AGAGCAAAAGGATGTTTACCCTCCGCGGATGGC-CTCGATCGTTGGGCTGTCTATCGC 1528
QY 1242 TAGCGGAAGAAATTAAGAGCAGCAATTCATTTGATACAAAATCTATTTAAAGTTACACC 1301
DB 1529 GCGCAAGGAATTCCTATGAAGTCCGAATCTACTTGACTCTTTGTATAGACAGCTAG 1588

```

QY 1302 AGCCATGGGCATGCAAAATGAGAAAAGC---AATAATGATGTAAGTGGATGACAGAACTGA 1358
 Db 1589 TGAATGGTCTTTAGAAATTCGAAGCCGCCAGGAATTCATAATTTATGATGATGCGCACTGG 1648
 QY 1359 AGCCTACTTAAGAGCTTACAGCAAAAGTTCACAGACACCCAGATAGTTGTCTGTCT 1418
 Db 1649 AACTTATGTGAGAGCAATGGATGATGTGTGGCTTCAGATCCCAACTTATATATGCT 1708
 QY 1419 GTTGTCAAGTAATCGGAAGGCAAAATACGATCTATTAAAAATACCTGTGTACAGATTG 1478
 Db 1709 CGTACCCCAATGATAACGCCGAAGATACTCATCAATCAAAAGAGAGAGATAGTTGACAG 1768
 QY 1479 CCTACCCCAAGTCACTGTGTGGTGGCCGGAACCTTAGTCAAAACAGCAAACTGTCAAGC 1538
 Db 1769 GCGGTGGCCAACTCA---AGTTGTGACCTTAAACAGCAGCAAGAACCGTAGCTTATGAG 1825
 QY 1539 CATTTGTACAAAGATTGCCCTACAGATGAACCTGCAAGATGGGAGAGAGCTCTGGAGGT 1598
 Db 1826 CATTTGCCCAAAATAGCAATCCAACTGAATGCAAGTTGGGATATACACCTTGGATGAT 1885
 QY 1599 GGACATCCCTTGAAGCTGTGTGATGATCGTTGGCATCGATTGTACCATGACATGACAGC 1658
 Db 1886 CGAATCACTCCCTGTCCGACTGATGACAAATTTGCTTTCATGCTGCAAGAGCAGACAGAGA 1945
 QY 1659 TGGCGGAGTCAATCGCAGGATTTGTTGCCAGCATCAATGAAGGGATGACCCGCTGGTT 1718
 Db 1946 TCGGAAGAGGGCCTACCGGAGCATTTGATTCCTCAATGGATCTACAGCAAAACTCCACGTA 2005
 QY 1719 CTCAGCTGTCATATTTTCAGGATAGAGGACAGAGCTGGTGTAGATGGCTCAAGTCTGGCT 1778
 Db 2006 CTTGAGCAGCTCAGGAGTGCAGCGCTTTGATGTGCTCGCTAAACCCCTTTGGCCGAT 2065
 QY 1779 GCAAGC---GGCTCTGAGGGCTTGAATAGCTGCAATGAGTACATGCCAGCGGATCAT 1835
 Db 2066 GATACAAAGGCCCTGCGCCATATCAACATGAGCATAGGAAGCTGCCATCTGATCGT 2125
 QY 1836 CGTGTACCGGATGCGTAGAGAGCGGCAGCTGAAACACATGTGTGAACACTGAGAGTCC 1895
 Db 2126 ATTTTATCGACGCGTGTGAGCTCGGCTCTCTAAAGCAGCTTTTGAATTTGAAGTCAA 2185
 QY 1896 ACAGTTTTTGGATTCTTAAATCCATTGGTAGAGTTTACACCTTACACTAAACGGTAAT 1955
 Db 2186 GGAATCATTTGAGAGTTGAAATCAATGAAATACGCCGCGCTCAGCTTAAGCCAGCGCAAT 2245
 QY 1956 TGTGTGAGAAAAGAGTGAACACAGATTTTGTCTCAGTCTGAGGAGAGACTTCAGAA 2015
 Db 2246 AGCT---TATATTGTGTGTAACAGATCCATGAACACGCGCTTCTTCTCAACGACAAA 2302
 QY 2016 TCCATCTCTGAAAGTATTGTATGATGAGGTTTACAGAGTGTACAGCAGATGTTGACTTTT 2075
 Db 2303 TCCCTCCGCTGGTACTATAGTTGATGAGCTTATTAACCTGCGCCGAGATACGACTTTTA 2362
 QY 2076 TATCGTAGCCAGGCTGTGAGAAGTGTAGTGTCTTCCACACATTAACAATGTCATCTA 2135
 Db 2363 TCTGCTCTGCAACAAGTTGCTGAGGTACAGTGTGCGCCAGCAGCTACATGTTCTTTA 2422
 QY 2136 TGACAAAGCGGCTGAAGCCAGACCAATACAGCGCTTGAACCTTACAGCTGTGCGCAAT 2195
 Db 2423 TAGCAGCATGGTGTCTCACCAGGAGAAATGCAAAAACCTTACGTACAGATGTGCACTT 2482
 QY 2196 CTATTACAACCTGGCAGGTGTCATGCTGTTCTCTCTTGGCAGTACGCCACAGCT 2255
 Db 2483 GTACTACAATTTGTGGGCAACACAGAGTGGCAGAGCTTTGCGAGTACGTTAAGAAGCT 2542
 QY 2256 GGCTTTTCTTGTGGCCAGAGTATTTCAGAGAGCGCAATCTGTCACTGTCAAAACCGCT 2315
 Db 2543 AGCTACCTCTGCGGTAGCACTTGCATCTTATTCGCAAAACGCGCTCGAAGAAAGATT 2602
 QY 2316 TTACTACCTTAA 2328
 Db 2603 TTATTATCTATAA 2615

RESULT 14
 AAH14032
 ID AAH14032 standard; cDNA; 1733 BP.
 XX AAH14032;
 AC AAH14032;
 XX 26-JUN-2001 (first entry)
 XX Human cDNA sequence SEQ ID NO:11143.
 DE Human; primer; detection; diagnosis; antisense therapy; gene therapy; ss.
 XX Homo sapiens.
 XX EP1074617-A2.
 PN 07-FEB-2001.
 PD 28-JUL-2000; 2000EP-0116126.
 XX 29-JUL-1999; 99JP-0248036.
 PR 27-AUG-1999; 99JP-0300253.
 PR 11-JAN-2000; 2000JP-0118776.
 PR 02-MAY-2000; 2000JP-0183767.
 PR 09-JUN-2000; 2000JP-0241899.
 XX (HELI-) HELIX RES INST.
 XX Ota T, Isogai T, Nishikawa T, Hayashi K, Saito K, Yamamoto J;
 PI Ishii S, Sugiyama T, Wakamatsu A, Nagai K, Otsuki T;
 DR WPI; 2001-318749/34.
 XX Primer sets for synthesizing polynucleotides, particularly the 5602
 full-length cDNAs defined in the specification, and for the detection
 and/or diagnosis of the abnormality of the proteins encoded by the
 full-length cDNAs -
 PS Claim 8; SEQ ID 11143; 2537pp + CD ROM; English.
 CC The present invention describes primer sets for synthesizing 5602
 full-length cDNAs defined in the specification. Where a primer set
 comprises: (a) an oligo-dr primer and an oligonucleotide complementary
 to the complementary strand of a polynucleotide which comprises one of
 the 5602 nucleotide sequences defined in the specification, where the
 oligonucleotide comprises at least 15 nucleotides; or (b) a combination
 of an oligonucleotide comprising a sequence complementary to the
 complementary strand of a polynucleotide which comprises a 5'-end
 sequence and an oligonucleotide comprising a sequence complementary to a
 polynucleotide which comprises a 3'-end sequence, where the
 oligonucleotide comprises at least 15 nucleotides and the combination of
 the 5'-end sequence/3'-end sequence is selected from those defined in
 the specification. The primer sets can be used in antisense therapy and
 particularly full-length cDNAs. The primers are also useful for the
 detection and/or diagnosis of the abnormality of the proteins encoded by
 the full-length cDNAs. The primers allow obtaining of the full-length
 cDNAs easily without any specialised methods. AAH03166 to AAH13628 and
 AAH13633 to AAH18742 represent human cDNA sequences; AAH92446 to
 AAH95893 represent human amino acid sequences; and AAH13629 to AAH13632
 represent oligonucleotides, all of which are used in the exemplification
 of the present invention.
 SQ Sequence 1733 BP; 435 A; 408 C; 451 G; 439 T; 0 other;

Query Match 10.2%; Score 237.2; DB 22; Length 1733;
 Best Local Similarity 52.1%; Pred. No. 9.3e-60;
 Matches 608; Conservative 0; Mismatches 548; Indels 12; Gaps 3;
 QY 1170 AGATTGGTCCAAAGAACACAGAGTGCACCATTAATAGTGTAAAGCCACTAGATACTG 1229
 Db 3 AAATGGTGTAGGAAGTACACGAGACCCCTTCCATCTTGACTATCCCATGATTCTG 62


```
Qy 1230 GCTGTGATCTATAGCGAAGAAATTTATGACGACCAATTCATTGATACAAATCTATT 1289
Db 63 GGCACCTTTTACCACAGAGACATGACGAGCTCGAAGCTGGTCAACATGTTGGA 122
Qy 1290 TAAAGTTACACGAGCATGGGATGCAATGAGAAAGCAATTAATGATTGAAGT---GGA 1346
Db 123 GAAGATAGCGGCCCCATTTGGCATGCGTATGAGCCACCGGCTGGTTGAACAAAGGA 182
Qy 1347 TGACAGACTGAGGCTTACTTAAGAGTCTTACGCAAAAGGTCAAGAG-----ACAC 1400
Db 183 TGACCGAATAGAGACTTATGTGAGAACCAATTCATCCAGCTTAGGAGCTGAGGGAAGAT 242
Qy 1401 CCAGATAGTGTCTGTCTGTGTTGTCAGTAATCGGAAGGCAAAATACGATGCTATTAAAAA 1460
Db 243 ACAGATGTTGTTGTCATCATCATGGGCCACGATGATGATCTCATGCGGCATCAATAA 302
Qy 1461 ATACCTGTGTACAGATGCCCTACCCCAAGTCAGTGTGTGTGGCCCAACCTTAGGCAA 1520
Db 303 GCTGTGCTGTGTGACAGTCCCAAGTCCCTCCAGGTGTGTCATGTTGCAACCAATTGGTCA 362
Qy 1521 ACAGCAACTGTCATGGCCATTTGCTACAAAGNTTGCCTACAGATGAAGTCAAGATGGG 1580
Db 363 GCCCACAGGCTTCGAGTGTGGCCACGAGAAATTTTACTTCAGATTAATGTAAATGGG 422
Qy 1581 AGGAGAGCTCTGGAGGTGGACATCCCTCTGAAGCTCGTGTGATGATCGTTGGCATCGATTG 1640
Db 423 TGGTGAGCTCTGGGAGTGGATATTCCTCTGAACAGTTAATGGTGATCGGGATGATGT 482
Qy 1641 TTACATGACATGACAGTGGCGGAGGTCAATCGCAGAGTTGTTGGCAGCATCAATGA 1700
Db 483 TTACCATGACCCAGTAGAGGATCGCTCGTGTGTTGGCTTCGTGGCAAGCATCAATCT 542
Qy 1701 AGGATGACCCGCTGTTCTACGCTGCATATTTCCAGNTAGAGACAGGAGCTGTGAGA 1760
Db 543 CACCTTCAAAATGGTATTCGGGTGGTGTTCAGATGCCGATCAGGAGATGTTGGA 602
Qy 1761 TGGGCTCAAAAGTCTGCTGCAAGCGCTCTGAGGCTTGGATAGCTGCAATGAGTACAT 1820
Db 603 CAGCTGAGCTATGCTCTCGTGGCTCTTAAAGTTTATGAGTGACCACTGTCT 662
Qy 1821 GCCCAGCCGATCATCTGTGTCGCGATGGCGTAGGACGCGCAGCTGAAACACTGGT 1880
Db 663 ACCAGAGAAGATTTGGTGTACCGGTGAGTGTCTGATGCCAACTGAAGACAGTTGC 722
Qy 1881 GAATACGAGTGCACAGTTTGGATGCTTAAATCCATTTGTAAGGTTACAACCC 1940
Db 723 CACTATGAGATCTCTCAACTACAGAAAGTGTGAAAGCTTTTGAGA---ATTATCAGCC 779
Qy 1941 TAGACTAACGGTAATTTGGTGCAAGAAAGAGTGAACACCAAGATTTTGTCTAGTCTGG 2000
Db 780 CAAGATGGTGTGTTGTAGTTCAGAGAAATCAGTACTAATCTATCTGCTGCTGCC 839
Qy 2001 AGGAGACTTCAGATCCACTCTCTGGAACAGTTATGATGAGGTTACAGACAGA 2060
Db 840 TCAGAACTTTGTAATCCCACTCTCTGGAACAGTGTGTAGATCATACAAATAAGCTGGA 899
Qy 2061 ATGTATGACTTTTATTCGAGCCAGCGCTGTGAGAGTGTGAGTGTCTTCTCCACACA 2120
Db 900 GTGGGTGATTTTATCTTCTTGGCCATCATGTACGGCAGGCGCTGTGCACTTCTACGCA 959
Qy 2121 TTACAATGTCTATGACAAACAGCGGCTTGAAGCCAGACACATACAGCGCTTGAACCTA 2180
Db 960 TTATGTCTGTGTTCTCAACACGCAAACTGAGCCCTGATCATATGAGAGGCTGACTTT 1019
Qy 2181 CAAGTGTGCCATCTATTAACTGCCAGGTGTCAATCGTGTTCCTGCTCTCTGCCA 2240
Db 1020 CAAACTGTGCCACATGACTGGAATTGGCTGGCACCATCAGAGTTCAGCTCTCTGCAA 1079
Qy 2241 GTAGCCCAAGCTGGCTTTTCTTGTGGCCAGAGTATTCAGAGAGGCCAAATCTGTC 2300
Db 1080 GTATGCCCAAGTAGCTTTTCTGTGTCAGSACATCTTGTGATCATGAAACGACCATCCA 1139
```

```
Qy 2301 ACTGTCAAACCGCCTTTTACTACTCTAA 2328
Db 1140 GCTGTGGAGAACCTGTTCTTCTCTGTGA 1167

RESULT 15
ABZ19102
ID ABZ19102 standard; cDNA; 209 BP.
XX AC ABZ19102;
XX DT 23-JAN-2003 (first entry)
XX Group III cDNA cancer related clone SEQ ID NO:1528.
XX Human; cancer; tumour; therapy; diagnosis; CT antigen; CP antigen;
XX immune response; virology; immunology; microbiology; molecular biology;
XX recombinant DNA technology; gene; ss.
XX Homo sapiens.
XX WO200278516-A2.
XX PD 10-OCT-2002.
XX PF 28-MAR-2002; 2002WO-US10421.
XX PR 30-MAR-2001; 2001US-280255P.
XX PR 28-AUG-2001; 2001US-315563P.
XX PR 09-JAN-2002; 2002US-347313P.
XX PA (CORI-) CORIXA CORP.
XX PI Wang T, Wang S, Bangur CS, Gaiger A;
XX WIPI; 2003-058387/05.
XX New immunogenic polynucleotides or polypeptides useful for diagnosing,
XX preventing and treating cancer expressing CT or CP mRNA antigens, and
XX in virology, immunology, microbiology, molecular biology and
XX recombinant DNA techniques
XX Claim 1; SEQ ID 1528; 207pp; English.
XX AB017575 to AB020506 represent isolated polynucleotide (I) sequences, and
XX ABP54446 to ABP54472 represent protein (II) sequences, from the present
XX invention. (I) and (II) have cytostatic activity and can be used in gene
XX therapy and vaccines. (I), (II), antibodies and compositions from the
XX present invention are useful for diagnosing, preventing and treating
XX cancer, which expresses CT or CP mRNA antigens. They are useful for
XX stimulating immune response. They can also be useful in virology,
XX immunology, microbiology, molecular biology and recombinant DNA
XX techniques.
XX N.B. The sequence data for this patent did not form part of the printed
XX specification, but was obtained in electronic format directly from WIPO
XX at ftp.wipo.int/pub/published_pot_sequences.
XX SQ Sequence 209 BP; 48 A; 47 C; 70 G; 44 T; 0 other;

Query Match 8.9%; Score 207.4; DB 25; Length 209;
Best Local Similarity 99.5%; Pred. No. 2e-51;
Matches 208; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1658 CTGGGGGAGGTCAATGCGCAGGATTTTGTCCAGCATCAATGAAGGATACCGCGTGGT 1717
Db 1 CTGGGGGAGGTCAATGCGCAGGATTTTGTCCAGCATCAATGAAGGATACCGCGTGGT 60
Qy 1718 TCTCAGCTGCATATTTCCAGATAGAGGACAGAGCTGGTAGATGGCTCAAGTCTGGC 1777
Db 61 TCTCAGCTGCATATTTCCAGATAGAGGACAGAGCTGGTAGATGGCTCAAGTCTGGC 120
Qy 1778 TGCAAGCGGCTCTGAGGCTTGGATAGCTGCAATGATGATGTCATGCCAGCGGATCATCG 1837
```


D0 121 TGCAGGGGCTCTGAGGGCTTGGATAGCTGCAATGAGTACATGCCCGCCGGATCATCG 180

QY 1338 TGTACCGCGATGGCGTAGGAGACGGCCAG 1866

Db 181 TGTACCGCGATGGCGTAGGAGACGGCCAG 209

Search completed: December 6, 2003, 12:06:50
Job time : 624 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: December 6, 2003, 11:30:43 ; Search time 4855 Seconds
(without alignments)
11654.134 Million cell updates/sec

Title: US-10-043-774B-1

Perfect score: 2328

Sequence: 1 atgatcttgggtggaacac.....acggccttactacctctaa 2328

Scoring table: IDENTITY_NUC

Gapop 10.0 , Gapext 1.0

Searched: 22781392 seqs, 12152238056 residues

Total number of hits satisfying chosen parameters: 45562784

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

EST:*
1: em_estba:*
2: em_esthum:*
3: em_estin:*
4: em_estmu:*
5: em_estov:*
6: em_estpl:*
7: em_estro:*
8: em_htc:*
9: gb_est1:*
10: gb_est2:*
11: gb_htc:*
12: gb_est3:*
13: gb_est4:*
14: gb_est5:*
15: em_estfun:*
16: em_estom:*
17: em_gss_hum:*
18: em_gss_inv:*
19: em_gss_pln:*
20: em_gss_vrt:*
21: em_gss_fun:*
22: em_gss_mam:*
23: em_gss_mus:*
24: em_gss_pro:*
25: em_gss_rod:*
26: em_gss_phg:*
27: em_gss_vrl:*
28: gb_gss1:*
29: gb_gss2:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	575.6	24.7	874	10	BF666315
2	515.2	22.1	577	10	BF080918
3	513.4	22.1	527	12	BM751055
4	502.4	21.6	1495	10	BF664941

5	492	21.1	705	10	BF246708
6	471.2	20.2	533	10	BF080923
7	459	19.7	3559	11	AK030116
8	441.4	19.0	462	9	AA969938
9	438.8	18.8	818	12	EG773137
10	433.2	18.6	642	13	BU854142
11	415.2	17.8	565	10	BG101107
12	413	17.7	413	9	AL041473
13	405.6	17.4	779	12	BI560710
14	400.4	17.2	664	13	BU234973
15	390.6	16.8	1092	12	BI463482
16	388.8	16.7	420	13	BQ374910
17	387	16.6	587	9	AA430311
18	373	16.0	935	10	BF247005
19	370.8	15.9	608	13	BU230168
20	365.4	15.7	767	13	BU232120
21	350.6	15.1	875	14	CA471292
22	347	14.9	882	12	BI459464
23	324.6	13.9	715	10	BG721550
24	307.8	13.2	665	12	BI463897
25	305.8	13.1	622	9	AL705072
26	301	12.9	731	12	BI458664
27	296.8	12.7	640	9	AL705301
28	296	12.7	833	13	BU275338
29	285.6	12.3	767	14	CB592670
30	285.2	12.3	609	13	BQ481304
31	275.8	11.8	559	12	BM728018
32	272	11.7	807	13	BU160046
33	265	11.4	617	12	BJ040023
34	264.8	11.4	512	12	BM680592
35	258.4	11.1	477	12	BM539361
36	249	10.7	718	13	BQ435882
37	245.2	10.5	568	13	BQ074332
38	243.4	10.4	785	10	BG221179
39	242.2	10.4	378	10	BF764928
40	241.2	10.4	534	12	BI672860
41	241.2	10.4	800	10	BG206124
42	240.2	10.3	796	10	BG209773
43	238.2	10.2	778	10	BG210801
44	238.2	10.2	783	10	BG198827
45	238	10.2	795	10	BG191694

ALIGNMENTS

RESULT 1
BF666315
LOCUS
DEFINITION 602119233F1 NIH_MGC_56 Homo sapiens cDNA clone IMAGE:4276136 5',
874 bp mRNA linear EST 21-DEC-2000
mRNA sequence.
ACCESSION BF666315
VERSION BF666315.1 GI:11940210
KEYWORDS EST.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 874)
NIH-MGC <http://mgc.nci.nih.gov/>.
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgabbs-remail.nih.gov
Tissue Procurement: ATCC
cDNA Library Preparation: CLONTECH Laboratories, Inc.
DNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at:
<http://image.llnl.gov>
Plate: LLCMI096 row: b column: 09
High quality sequence stop: 594.

FEATURES
source

Location/Qualifiers

```
1..874
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="IMAGE:4276136"
/tissue_type="primitive neuroectoderm"
/lab_host="DH10B (T1 phage-resistant)"
/clone_lib="NIH_MGC_56"
/Note="Organ: Brain; Vector: pDNR-LIB (Clontech); Site_1:
SfiI (ggccctggcc); Site_2: SfiI (ggccattggcc);
Double-stranded cDNA was prepared from cell line RNA. 5'
and 3' adaptors were used in cloning as follows: 5'
adaptor sequence: 5'-ATCTAGAGCGGCGGCGGACATG-dr(30)BN-3'
(where B = A, C, or G and N = A, C, G, or T). Average
insert size 1.65 kb (range 0.9-4.0 kb). 15/15 colonies
contained inserts by PCR. This library was enriched for
full-length clones and was constructed by Clontech
Laboratories (Palo Alto, CA)."
BASE COUNT 200 a 218 c 258 g 198 t
ORIGIN
Query Match 24.7%; Score 575.6; DB 10; Length 874;
Best Local Similarity 94.0%; Pred. No. 1.5e-142;
Matches 654; Conservative 0; Mismatches 34; Indels 8; Gaps 5;
QY 1333 ATGATTGAAGTGGATGACAGAACTGAAGCTTACTTAAGAGTCTTACACAAAGGTGACA 1392
Db 8 AGGATTGAAGTGGATGACAGAACTGAAGCTTACTTAAGAGTCTTACACAAAGGTGACA 67
QY 1393 CGAGACACCCAGATAGTGTCTGTCTGTGTTCACAGTAATCGGAGGACAAATACGATGCT 1452
Db 68 CGAGACACCCAGATAGTGTCTGTCTGTGTTCACAGTAATCGGAGGACAAATACGATGCT 127
QY 1453 ATTAATAAATACCTGTGTACAGATGTCCTACCCCAAGTCAAGTGTGTGGTGGCCGAC 1512
Db 128 ATTAATAAATACCTGTGTACAGATGTCCTACCCCAAGTCAAGTGTGTGGTGGCCGAC 187
QY 1513 TTAGCAACACGAACTGTGTATGATGTCATGATGTCATGATGTCATGATGTCATGATGTC 1572
Db 188 TTAGCAACACGAACTGTGTATGATGTCATGATGTCATGATGTCATGATGTCATGATGTC 247
QY 1573 AAGTGGGAGGAGAGCTGTGGAGGTGTGACATCCCTTGAAGCTGTGATGATGTTGGC 1632
Db 248 AAGTGGGAGGAGAGCTGTGGAGGTGTGACATCCCTTGAAGCTGTGATGATGTTGGC 307
QY 1633 ATCGATTGTTTACCATGACATGACAGCTGGGGAGGTCAATCGCAGGATTTGTTGCCAGC 1692
Db 308 ATCGATTGTTTACCATGACATGACAGCTGGGGAGGTCAATCGCAGGATTTGTTGCCAGC 367
QY 1693 ATCAATGAAGGATGACCCGCTGTTCTCAGCTGTCATATTTTCAGATPAGAGGACAGGAG 1752
Db 368 ATCAATGAAGGATGACCCGCTGTTCTCAGCTGTCATATTTTCAGATPAGAGGACAGGAG 427
QY 1753 CTGGTAGATGGCTCAAAAGTCTGCTGCAAGCGCTCTGAGGGCTTGAATAGCTGCAAT 1812
Db 428 CTGGTAGATGGCTCAAAAGTCTGCTGCAAGCGG-TCTGAGGGCTTGAATAGCTGCAAT 486
QY 1813 GAGTACATGCCAGCGGATCATGCTGTACCGCGATGGCGTAGGACGGCGCATGCTGAAA 1872
Db 487 GAGTACATGCCAGCGGATCATGCTGTACCGCGATGGCGTAGGACGGCG--AGTGAAA 544
QY 1873 ACACCTGTGAATACGAAAGTCCAGATTTTGGATGTCCTTGAATGTCCTTGAATGTCCTTGA 1932
Db 545 ACACCTGTGAATACGAAAGTCCAGAG-TTTTGGATGTCCTTGAATGTCCTTGAATGTCCTTGA 603
QY 1933 TACACCTGTGAATACGAAAGTCCAGATTTTGGATGTCCTTGAATGTCCTTGAATGTCCTTGA 1992
Db 604 T-CACCTGTGAATACGAAAGTCCAGAG-TTTTGGATGTCCTTGAATGTCCTTGAATGTCCTTGA 659
QY 1993 CAGTCTGAGGAGAGACTTCAGAAATCCACTTCTCTGGA 2028
```

Db

660 CTCGCCCGGAGAGACTTCGGAGACATCCTTCTCTGGA 695

RESULT 2

```
BF080918
LOCUS BF080918 577 bp mRNA linear EST 18-OCT-2000
DEFINITION 233684 MARC 2P1G Sus scrofa cDNA 5', mRNA sequence.
ACCESSION BF080918
VERSION BF080918.1 GI:10874739
KEYWORDS EST.
SOURCE Sus scrofa (pig)
ORGANISM Sus scrofa
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Cetartiodactyla; Suina; Suidae; Sus.
REFERENCE 1 (bases 1 to 577)
AUTHORS Fahrenkrug,S.C., Smith,T.P.L., Feking,B.A., Cho,J., White,J.,
Vallet,J., Wise,T., Rohrer,G.A., Perte,G., Sultana,R., Quackenbush
J. and Keele,J.W.
TITLE Porcine gene discovery by normalized cDNA-library sequencing and
EST cluster assembly
JOURNAL Mamm. Genome 13 (8), 475-478 (2002)
MEDLINE 12226715
PUBMED 12226715
COMMENT Contact: Smith TPL
USDA, ARS, US Meat Animal Research Center
PO Box 166, Clay Center, NE 68933-0166, USA
Tel: 402 762 4366
Fax: 402 762 4390
Email: smith@email.marc.usda.gov
Single pass sequencing. Bases called and alt trimmed with phred
v0.980904.e. Vector identified by cross_match with the -minscore 18
and -mismatch 12 options.
PCR Primers
FORWARD: AGGAAACAGCTATGACCAT
BACKWARD: GTTTCACAGTCACGACG
Plate: 53 row: N column: 21
Seq primer: ATTGAGTGACACTATAG.
Location/Qualifiers
1..577
/organism="Sus scrofa"
/mol_type="mRNA"
/db_xref="taxon:9823"
/tissue_type="pooled"
/lab_host="DH10B"
/clone_lib="MARC 2P1G"
/Note="Vector: pCMV SPORT6; Site_1: NotI; Site_2: SalI;
Library made from pooled tissue from testis, ovary,
endometrium, hypothalamus, pituitary, and placenta."
BASE COUNT 198 a 98 c 132 g 149 t
ORIGIN
Query Match 22.1%; Score 515.2; DB 10; Length 577;
Best Local Similarity 93.4%; Pred. No. 1.8e-126;
Matches 538; Conservative 0; Mismatches 38; Indels 0; Gaps 0;
QY 1132 CAAGTGGGAAAAACATTTGATTACAATCCCAATTTGCAGATTGGTCCAAAGAAACAAGA 1191
Db 2 CAAGTGGGAAAAACATTTGATTACAATCCCAATTTGCAGATTGGTCCAAAGAAACAAGA 61
QY 1192 GGTGACCATTAATTAGTGTTAAGCCACTAGATACTGGCTGTGTATCTATACGGAAGA 1251
Db 62 GGTGACCATTAATTAGTGTTAAGCCACTAGATACTGGCTGTGTATCTATACGGAAGA 121
QY 1252 AATTATGAAGCAGCCCAATTCATTGTATACAAAATCTATTTAAAGTTACACGAGCATGGGC 1311
Db 122 AATTATGAAGCAGCCCAATTCATTGTATACAAAATCTATTTAAAGTTACACGAGCATGGGC 181
QY 1312 ATCCAAATGAAAAAGCAATTAATGATTGAAGTGGATGACAGAACTGAAGCCTACTTAAGA 1371
Db 182 ATCAAAATGAAAAAGCAATTAATGATTGAAGTGGATGACAGAACTGAAGCCTACTTAAGA 241
QY 1372 GTCTTACAGAAAAGTTCACGACAGACCCAGATAGTGTGTGTGTGTGTCAAGTAAT 1431
```

```

FEATURES
source
high quality sequence stop: 327.
Location/Qualifiers
1. 527
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="S9SNU601-21-E08"
/sex="M"
/tissue_type="Ascites"
/cell_type="Epithelial"
/cell_line="SNU-601"
/lab_host="Top10F'"
/clone_lib="S9SNU601"
/note="Organ: Stomach; Vector: pME18-FL3; Site 1: XhoI;
Site 2: XhoI; The poly (A) + RNA was dephosphorylated with
bacterial alkaline phosphatase (BAP) and then decapped
with tobacco acid pyrophosphatase (TAP). The decapped
intact mRNA was ligated with DNA-RNA linker including SfiI
site by treatment of T4 RNA ligase and the first strand
cDNA was synthesized with Superscript II using SfiI
oligo-dT primer. After first strand synthesis, RNA was
degraded by NaOH treatment and cDNA was amplified by PCR
reaction. The PCR products were digested with SfiI and
cloned into DraIII-digested pME18-FL3 vector. The

```

found through the I.M.A.S.E. Consortium/HEALE at:
<http://image.llnl.gov>

<http://image.llnl.gov>

Plate: LLCM1096 ROW: f COLUMN: 24

High quality sequence stop: 511.

FEATURES

Location/Qualifiers

1..1495

/organism="Homo sapiens"

/mol_type="mRNA"

/db_xref="taxon:9606"

/clone="IMAGE:4276247"

/tissue_type="primitive neuroectoderm"

/lab_host="DH10B (T1 phage-resistant)"

/clone_lib="NIH MGC 56"

/note="Organ: brain; Vector: pDNR-LIB (Clontech); Site_1:"

SfiI (ggccattggcc); Site_2: SfiI (ggccattggcc);

Double-stranded cDNA was prepared from cell line RNA. 5'

and 3' adaptors were used in cloning as follows: 5'

adaptor sequence: 5'-CAGGCCATTATGGCC-3' and 3' adaptor

sequence: 5'-ATTCTAGAGCGCGCGGCACATG-dt(30)BN-3'

(where B = A, C, G and N = A, C, G, or T). Average

insert size 1.65 kb (range 0.9-4.0 kb). 15/15 colonies

contained inserts by PCR. This library was enriched for

full-length clones and was constructed by Clontech

Laboratories (Palo Alto, CA)."

BASE COUNT 476 a 280 c 536 g 203 t

ORIGIN

Query Match 21.6%; Score 502.4; DB 10; Length 1495;

Best Local Similarity 94.7%; Pred. No. 8.4e-123;

Matches 553; Conservative 0; Mismatches 26; Indels 5; Gaps 3;

QY 1333 ATGATTGAAGTGGATGACAGAACTCAAGCTACTTAAGAGTCTTACAGCAAAAGGTGACA 1392

DB 8 AGGATTGAAGTGGATGACAGAACTCAAGCTACTTAAGAGTCTTACAGCAAAAGGTGACA 67

QY 1393 GCAGACACCAGATAGTGTCTGTCTGTCTCAAGTAATCGAAGGACAAATACGATGCT 1452

DB 68 GCAGACACCAGATAGTGTCTGTCTGTCTCAAGTAATCGAAGGACAAATACGATGCT 127

QY 1453 ATTAAAAATACCTCTGTACAGATTCCTTACCACCAAGTCAGTGTGTGGTGGCCGACACC 1512

DB 128 ATTAAAAATACCTCTGTACAGATTCCTTACCACCAAGTCAGTGTGTGGTGGCCGACACC 187

QY 1513 TTAGGCAACAGCAAACTGTCTATGCCATTTGTCTCAAGAGTTGCCCTACAGATGAAGTCG 1572

DB 188 TTAGGCAACAGCAAACTGTCTATGCCATTTGTCTCAAGAGTTGCCCTACAGATGAAGTCG 247

QY 1573 AAGATGGAGGAGAGCTCTGGAGGTGACATCCCTTGAAGCTCGTGATGATCGTTGGC 1632

DB 248 AAGATGGAGGAGAGCTCTGGAGGTGACATCCCTTGAAGCTCGTGATGATCGTTGGC 307

QY 1633 ATCGATTCTTACCATGACATGACAGCTGGCGGAGTCAATCGCAGGATTTGTTCGACAGC 1692

DB 308 ATCGATTCTTACCATGACATGACAGCTGGCGGAGTCAATCGCAGGATTTGTTCGACAGC 367

QY 1693 ATCAATGAAGGATGACCCCGTGTCTTCAGCGTGCATATTTTCAGGATGAGGACAGGAG 1752

DB 368 ATCAATGAAGGATGACCCCGTGTCTTCAGCGTGCATATTTTCAGGATGAGGACAGGAG 427

QY 1753 CTGTAGATGGCTCAAGATCTGCTGCAAGCGCTCTGAGGGCTTGAATGAGCTGCAAT 1812

DB 428 CTGTAGATGGCTCAAGATCTGCTGCAAGCGG-TCTGAGGGCTTGAATGAGCTGCAAT 486

QY 1813 GAGTACATGCCAGCCGATCATCG-TCTACCGGATGGCTTAGGACGCGCCAGCTGAA 1871

DB 487 GAGTACATGCCAGCCGATCATCGTTTACCGGATGGGTTAGAGAGGCACTGT--- 543

QY 1872 AACACTGTGTAACACGAAGTCCACAGCTTTTGGATTGCTCTAA 1915

DB 544 AACCCGCTGGACTACGAGGCGGCCAGCTTTTGGTTCCTTAA 587

RESULT 5

BF246708

LOCUS

BF246708 705 bp mRNA linear EST 14-NOV-2000

DEFINITION

mRNA sequence.

ACCESSION

BF246708

VERSION

BF246708.1

KEYWORDS

EST

SOURCE

Homo sapiens (human)

ORGANISM

Homo sapiens

REFERENCE

1. (Bases 1 to 705)

AUTHORS

NIH-MGC <http://mgc.nci.nih.gov/>

TITLE

National Institutes of Health, Mammalian Gene Collection (MGC)

JOURNAL

UNPUBLISHED

COMMENT

Contact: Robert Strausberg, Ph.D.

Email: cgapbs-remail.nih.gov

Tissue Procurement: ATCC

cDNA Library Preparation: CLONETECH Laboratories, Inc.

DNA Sequencing by: Incyte Genomics, Inc.

Clone distribution: MGC clone distribution information can be

found through the I.M.A.G.E. Consortium/LLNL at:

<http://image.llnl.gov>

Plate: LLCM927 row: m column: 02

High quality sequence stop: 572.

Location/Qualifiers

1..705

/organism="Homo sapiens"

/mol_type="mRNA"

/db_xref="taxon:9606"

/tissue_type="glioblastoma"

/lab_host="DH10B (T1 phage-resistant)"

/clone_lib="NIH MGC 57"

/note="Organ: brain; Vector: pDNR-LIB (Clontech); Site_1:"

SfiI (ggcgcttgcc); Site_2: SfiI (ggcgcttgcc);

Double-stranded cDNA was prepared from cell line RNA. 5'

and 3' adaptors were used in cloning as follows: 5'

adaptor sequence: 5'-CAGGCCATTATGGCC-3' and 3' adaptor

sequence: 5'-ATTCTAGAGCGCGCGGCACATG-dt(30)BN-3'

(where B = A, C, G and N = A, C, G, or T). Average

insert size 1.55 kb (range 0.9-4.0 kb). 12/15 colonies

contained inserts by PCR. This library was enriched for

full-length clones and was constructed by Clontech

Laboratories (Palo Alto, CA)."

BASE COUNT 206 a 145 c 194 g 160 t

ORIGIN

Query Match 21.1%; Score 492; DB 10; Length 705;

Best Local Similarity 93.8%; Pred. No. 3.3e-120;

Matches 621; Conservative 0; Mismatches 30; Indels 11; Gaps 10;

QY 1333 ATGATTGAAGTGGATGACAGAACTCAAGCTACTTAAGAGTCTTACAGCAAAAGGTGACA 1392

DB 11 AGGATTGAAGTGGATGACAGAACTCAAGCTACTTAAGAGTCTTACAGCAAAAGGTGACA 70

QY 1393 GCAGACACCAGATAGTGTCTGTCTGTCTCAAGTAATCGAAGGACAAATACGATGCT 1452

DB 71 GCAGACACCAGATAGTGTCTGTCTGTCTCAAGTAATCGAAGGACAAATACGATGCT 129

QY 1453 ATTAAAAATACCTCTGTACAGATTCCTTACCACCAAGTCAGTGTGTGGTGGCCGACACC 1512

DB 130 ATTAAAAATACCTCTGTACAGATTCCTTACCACCAAGTCAGTGTGTGGTGGCCGACACC 189

QY 1513 TTAGGCAACAGCAAACTGTCTATGCCATTTGTCTCAAGAGTTGCCCTACAGATGAAGTCG 1572

DB 190 TTAGGCAACAGCAAACTGTCTATGCCATTTGTCTCAAGAGTTGCCCTACAGATGAAGTCG 249

QY 1573 AAGATGGAGGAGAGCTCTTGGAGGGTGGACATCCCTTGAAGCTCGTGATGATCGTTGGC 1632

DB 250 AAGATGGAGGAGAGCTCTTGGAGGGTGGACATCCCTTGAAGCTCGTGATGATCG-TGGC 308

QY 1633 ATCGATTCTTACCATGACATGACAGCT-GGGCGGAGTCAATCGCAGGA-TTTGTGTCGA 1690

```

Db 309 ATCGATGTTACCATGACATGACAGCTGGGGCGGAGGTCAATCGCAGGATTTTGTGCCA 368
Qy 1691 GCATCAATGAAGGGATGACCCGCTGTTCTCAGCTGCATATTTTCAGATAGACGACAGG 1750
Db 369 GCATCAATGAAGGGATGACCCGCT-GTTCTCAGCTGCATATTTTCAGATAGACGACAGG 427
Qy 1751 AGCTGGTAGATGGGCTCAAAGTCTCCCTGCAAGCGGCTCTGAGGCT-TGGATAGCT-- 1807
Db 428 AGCTGGTAGATGGGCTCAAAGTCTCCCTGCAAGCGGCTCTGAGGCTGTGGATAGCTTG 487
Qy 1808 GCAATGACTATACGCGACCGCGAT-CAATGCTGTCACCGGATGCGTGAAGAGCGCCAG 1866
Db 488 CAATGACTATACGCGACCGCGATCCATCGTGTACCGGATGCGTGAAGAGCGCCAG 547
Qy 1867 CTG-AAAACACTGGTGAAGTCTCGAAGTCCACAGTTTTCGATTCCTTAAATCC-ATTG 1924
Db 548 TTGAAAACACTGGTGAAGTCTCGAAGTCCACAGTTTTCGATTCCTTAAATCCAAATTG 607
Qy 1925 GTAGAGGTTACAAACCTAGACTTAACGGTAATTTGTGGAAGAAAAGAGTGAACACCAT 1984
Db 608 GTAGAGGTTACAAACCTAGAAAACGGTAATTTGTGGAAGAAAAGAGTGAACACCAT 667
Qy 1985 TT 1986
Db 668 TT 669

RESULT 6
BF080923
LOCUS
DEFINITION
ACCESSION
VERSION
KEYWORDS
SOURCE
ORGANISM
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Cetartiodactyla; Suina; Suidae; Sus.
1 (bases 1 to 533)
Fahrenkrug,S.C., Smith,T.P.L., Feking,B.A., Cho,J., White,J.,
Vallet,J., Wise,T., Rohrer,G.A., Pettea,G., Sultana,R., Quackenbush
,J. and Keefe,J.W.
Porcine gene discovery by normalized cDNA-library sequencing and
EST cluster assembly
Mamm. Genome 13 (8), 475-478 (2002)
22213789
12226715
Contact: Smith TPL
USDA, ARS, US Meat Animal Research Center
PO Box 166, Clay Center, NE 68933-0166, USA
Tel: 402 762 4366
Fax: 402 762 4390
Email: smith@email.marc.usda.gov
Single pass sequencing. Bases called and alt trimmed with phred
v0.980904.e. Vector identified by cross_match with the -minscore 18
and -minmatch 12 options.
PCR Primers
FORWARD: AGGAACACGCTATGACCAT
BACKWARD: GTTTCAGTCCAGCG
Plate: 53 row: O column: 21
Seq primer: ATTAGGTGACACTATAG.
FEATURES
Location/Qualifiers
1..533
/organism="Sus scrofa"
/mol_type="mRNA"
/db_xref="taxon:9823"
/tissue_type="pooled"
/lab_host="DH10B"
/clone_lib="MARC 2P1G"
/note="Vector: pCMV SPORT6; Site 1: NotI; Site 2: SalI;
Library made from pooled tissue from testis, ovary,
endometrium, hypothalamus, pituitary, and placenta."
BASE COUNT 185 a 93 c 119 g 136 t

```

ORIGIN

```

Query Match      20.2%; Score 471.2; DB 10; Length 533;
Best Local Similarity 92.9%; Pred. No. 1.1e-114;
Matches 494; Conservative 0; Mismatches 38; Indels 0; Gaps 0;

Qy 1132 CAAGGTGMAAAACATTTGATTACCAATCCCAATTCGAGATTGGTCCAAAGAACACAGA 1191
Db 2 CAAGGTGMAAAACATGGGATTACCAATCCCAATTCGAGATTGGTCCAAAGAACACAGA 61
Qy 1192 GGTGACCACTAATTAATAGTTTAAAGCCACTAGATACTGGCTGTGTGATCTATATACCGGAAGA 1251
Db 62 GGTGACCACTAATTAATAGTTTAAAGCCACTAGATACTGGCTGTGTGATCTATATACCGGAAGA 121
Qy 1252 AATTATGAAGCAGCAATTCATTGATCAAAATCTTATAAGTTTACACCGCATGGGC 1311
Db 122 AATTATGAAGCAGCAATTCATTGATCAAAATCTTATAAGTTTACACCGCATGGGC 181
Qy 1312 ATGCAAAATGAGAAAGCAATAATGATTGAAGTGTGATGACAGAACTGAAGCCTACTTAAAGA 1371
Db 182 ATACAAATGAAAAAGCAATAATGATTGAAGTGTGATGACAGAACTGAAGCCTACTTAAAGA 241
Qy 1372 GTCTTACAGAAAAGGTCACAGCAGACACCCAGATAGTTGTCTGTCTGTCTCAAGTAAT 1431
Db 242 GTCTTACAGAAAAGGTTACATCGATACCAGATAGTTGTCTGTCTGTCTCAAGTAAT 301
Qy 1432 CGAAGGACAAATACGATGCTTATTAATAAATACCTGTGTACAGATTGCCCTACCCCAAGT 1491
Db 302 CGAAGGACAAATATGATGCTTATTAATAAATACCTGTGTACAGATTGCCCTACCCCAAGT 361
Qy 1492 CAGTGTGTGGTGGCCCAACCTTAGGCAACACAGCAAACTGTCAATGGCCATTGTACAAAG 1551
Db 362 CAGTGTGTGGTGGCCCAACCTTAGGCAACACAGCAAACTGTCAATGGCCATTGTACAAAG 421
Qy 1552 ATTGCCCTACAGATGAATCAAGATCGGAGGAGCTCTGGAGGGTGGACATCCCCCTG 1611
Db 422 ATGCCCTGACAGTGAATCAAGATCGGAGGAGCTCTGGAGGGTGTATATGCCCTGA 481
Qy 1612 AAGTCTGTGATGCTGTGGATCGATTGTACATGACATGACAGCTGGGC 1663
Db 482 AAGTACGATGATGATGTGGCATTGTTATCATGATACACAGCTGGAC 533

```

RESULT 7

```

AK030116
LOCUS
DEFINITION
Mus musculus adult male testis cDNA, RIKEN full-length enriched
library, clone:4932443D15 product:piwi like homolog 1
(Drosophila)-like, full insert sequence.
AK030116
AK030116.1 GI:26326094
VERSION
HTC; CAP trapper.
KEYWORDS
Mus musculus (house mouse)
SOURCE
Mus musculus
ORGANISM
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
1
REFERENCE
Carninci,P. and Hayashizaki,Y.
High-efficiency full-length cDNA cloning
Meth. Enzymol. 303, 19-44 (1999)
9279253
PUBMED
10349636
2
REFERENCE
Carninci,P., Shibata,Y., Hayatsu,N., Sugahara,Y., Shibata,K.,
Itoh,M., Konno,H., Okazaki,Y., Muramatsu,M. and Hayashizaki,Y.
Normalization and subtraction of cap-trapper-selected cDNAs to
prepare full-length cDNA libraries for rapid discovery of new genes
Genome Res. 10 (10), 1617-1630 (2000)
20499374
PUBMED
11042159
3
REFERENCE
Shibata,K., Itoh,M., Aizawa,K., Nagaoka,S., Sasaki,N., Carninci,P.,
Konno,H., Akiyama,J., Nishi,K., Kutsunai,T., Tashiro,H., Itoh,M.,

```

Sumi, N., Ishii, Y., Nakamura, S., Hazama, M., Mishine, T., Harada, A., Yamamoto, R., Matsumoto, H., Sakaguchi, S., Ikegami, T., Kashiwagi, K., Fujiwaki, S., Inoue, K., Ogawa, Y., Izawa, M., Ohara, E., Watahiki, M., Yoneda, Y., Ishikawa, T., Ozawa, K., Tanaka, T., Matsuura, S., Kawai, J., Okazaki, Y., Muramatsu, M., Inoue, Y., Kira, A. and Hayashizaki, Y.
RIKEN integrated sequence analysis (RISA) system-384-format
sequencing pipeline with 384 multicapillary sequencer
Genome Res. 10 (11), 1757-1771 (2000)

TITLE

JOURNAL
MEDLINE
PUBMED
REFERENCE
AUTHORS

CDS

/strain="C57BL/6J"
/db_xref="PANTOM_DB:4932443D15"
/db_xref="taxon:10090"
/clone="4932443D15"
/sex="male"
/tissue_type="testis"
/clone_lib="RIKEN full-length enriched mouse cDNA library"
/dev_stage="adult"
142 - 3057
/note="unnamed protein product; piwi like homolog 1 (Drosophila)-like (MGD) MGI:1930036, GB|NM_021308, evidence: BLASTN, 99%, match=3559)
putative"
/codon_start=1
/protein_id="BAC26791.1"
/db_xref="GI:26326095"
/translation="MDPVRPLFRGTPVHPSCQVRMFCVQAPRPLPAMPGRAGPAG
RGLVFRPESSPLQPVQKDSVLGSMFRGMLDTAFRPFKREVPPLPGRVLRGL
SANVRKDRREPRSSLPDSVLAAGDSKLAASVGMRLGRSSEVSLPLGRVLRGL
IGRWKDPSPAFGLTARDPRLPQPPALSPSLHSDPPVLTWBERKEILLVKGSSK
GTQSLGLNLIKIOCHNEAVYQVHTVSPSVECKSMRGLKDHQSVGTGVTADPGSI
LPLVKLQVVELKSQKTDAAELSIKILQTKILEPCSDLCIPFNVVFRVRKLLDM
KLINGRYDPSAMVLOQRLLQIWPFGAASIRRTDGLGLLADLVSHKVRNLDVM
HATIQNKHFQDECSKLGVSIYITRYNNRTYRIDVDNNKTPKDSFMDGKEITF
LEYKNGYITVEDDQPLIHREPSQNNHMLLKGILLPLSFTMTGPIPERMKDL
FRAMKDTQINLSPKQHGALCCLQRIQSONEASNELTRWGLSLHDKVHKIGRL
EMERINLRNTSFVTSIDLNVKVEVTRDASILTIPMFHWPALFVPRKAMDQARELVNMLE
KIAGIPGRISPPAPVLEKDRIDETIYRTIQSLGVEGKIOMVCIIMGTDLLIGAI
KXLCVSPSPSQVINRTIGQPLRLSVAQKILLQNNKLGSLGVLGVDIPLKQMLVI
GVNHLPEKIVYRDSGDLQKTVANYEIPLOKQCEAFDNVHPKVVVVFVKKIST
NLIAPADPHFVTPSGTVDHTITSCWVDVFLLAHVROCGGPIPTHYICVLNTANLS
PDHMQRUTEKLCHEMWNWPGTIRVAPCKVAHKLAFLSGLLHHEPRLQICGNLFFL"
BASE COUNT 908 a 891 c 917 g 843 t

ORIGIN

Query Match 19.7%; Score 459; DB 11; Length 3559;
Best Local Similarity 52.4%; Pred. No. 5.8e-111;
Matches 1161; Conservative 0; Mismatches 1035; Indels 21; Gaps 6;
QY 124 GCCTATATCATGATCATCATGACTATAACCCACTGATGAAGCCAGCAAGACTCCGTCA 183
Db 850 GCAGTTTATCAGTACCATGAGCTTTTCAGCCCCAGTGTGAATCAAAAGCATGAGGTTT 909
QY 184 GCTCTCTTTTCAACACGAGAGTCTAATTTGGAAGTGTCTGCTTTTGTGGACCATTA 243
Db 910 GGCATGTTGAGGACCAACAGTCTGTCTGCTGGAACCTCACTGCTTTTGTGCTCTATT 969
QY 244 TTATTTTACCTAAAGACTACAGCAAAAGGTTACTGAAGTTTCTAGTAAGCCCGGAAT 303
Db 970 CTTTATCTCTCTGTTAAGCTTCAACAGTTGTTGAGTTAAAGTCAAGAGCAAGAACTGAC 1029
QY 304 GCAGAGATGTGAGGATAAGATCACTTTAACAATGAATCTTCCACTATCACTACCAACT 363
Db 1030 GATGCCGAGATCAGTATCAAGATTCAGTCAACAAGATCTCGAGCGCTGTCTGACCTG 1089
QY 364 TGTGTGAGTTCTATATATATTTTTCAGAGGCTTTTGAATCATGAAATTTGCACAA 423
Db 1090 TGCATCCCTTCTACAATGTTCTTCCGCGGGGTAAAGAACTTCGGAATATGAAGCTT 1149
QY 424 ATTGGAGAAATTTATATAACCAAAATGACCCAAATGATATCCAACTCAAGTTTGGTG 483
Db 1150 GTGGGAGAACTTCTATGACCTCAAGTGCCATGCTGACGCAACACAGATTGCAG 1209
QY 484 ATTTGGCTGCTTCACTACTTCCATCTTCAATGATGAACAGCATCATGCTGCTGCACT 543
Db 1210 ATCTGGCTGGCTATCGGCTAGTATCCGAGGACAGACGGGGTCTCTTCTCTGCTGCT 1269
QY 544 GAGCTTAGCCATAAAGTCCCTTGAAGTGAGACTGTTTGGATTTCATGTTCACTTTAT 603
Db 1270 GATGTTCTCTAAGGTCAITTCGGAACGACTCTGTGCTGATGATCATGCTATCTATC 1329
QY 604 CATCAGACAGACACATAAATTTCAAGAACAGATTTCACAAAGAACTAATAGCTTTAGTT 663

TITLE

JOURNAL
MEDLINE
PUBMED
REFERENCE
AUTHORS

The FANTOM Consortium and the RIKEN Genome Exploration Research Group Phase I & II Team
Analysis of the mouse transcriptome based on functional annotation of 60,770 full-length cDNAs
Nature 420, 563-573 (2002)

TITLE

JOURNAL
MEDLINE
PUBMED
REFERENCE
AUTHORS

6 (bases 1 to 3559)
Adachi, J., Aizawa, K., Akimura, T., Arakawa, T., Bono, H., Carninci, P., Fukuda, S., Furuno, M., Hanezaki, T., Hara, A., Hashizume, W., Hayashida, K., Hayatsu, N., Hiramoto, K., Hiraoka, T., Hirozane, T., Hori, F., Imocani, K., Ishii, Y., Itoh, M., Kagawa, I., Kasukawa, T., Kato, H., Kawai, J., Kojima, Y., Kondo, S., Kono, H., Kouda, M., Koyas, S., Kurihara, C., Matsuyama, T., Miyazaki, A., Murata, M., Nakamura, M., Nishi, K., Nomura, K., Numazaki, R., Ohno, M., Ohsato, N., Okazaki, Y., Saito, R., Saitoh, H., Sakai, C., Sakai, K., Sakazume, N., Sano, H., Sasaki, D., Shibata, K., Shinagawa, A., Shiraki, T., Sogabe, Y., Tegan, M., Tagawa, A., Takahashi, F., Takaku-Akahira, S., Takeda, Y., Tanaka, T., Tomaru, A., Toya, T., Yasunishi, A., Muramatsu, M. and Hayashizaki, Y.
Direct Submission

TITLE

JOURNAL
MEDLINE
PUBMED
REFERENCE
AUTHORS

Submitted (16-JUL-2001) Yoshihide Hayashizaki, The Institute of Physical and Chemical Research (RIKEN), Laboratory for Genome Exploration Research Group, RIKEN Genomic Sciences Center (GSC), RIKEN Yokohama Institute; 1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan (E-mail: genome-res@gsc.riken.go.jp, URL: http://genome.gsc.riken.go.jp/, Tel: 81-45-503-9222, Fax: 81-45-503-9216)

COMMENT

cDNA library was prepared and sequenced in Mouse Genome Encyclopedia Project of Genome Exploration Research Group in Riken Genomic Sciences Center and Genome Science Laboratory in RIKEN. Division of Experimental Animal Research in Riken contributed to prepare mouse tissues.
Please visit our web site for further details.
URL: http://genome.gsc.riken.go.jp/
URL: http://fantom.gsc.riken.go.jp/.

FEATURES

source

location/Qualifiers
1. .3559
/organism="Mus musculus"
/mol_type="mRNA"

Db	1330	CAGCAGAACAGGAGC---ACITTCAGGACGAGTGCAGCAAGCTTCTGGTTGGCAGCAAT	1386
Qy	664	GTCTTACCAGATATAACAATAAGACATACAGAGTGGATGATATTCAGCTGGGACAGAAAT	723
Db	1387	GTATACGCGGTACAACAATCGTACTCAGAAATCGATGATGTGATCTGGACACAGACC	1446
Qy	724	CCCAAGAGCACCTTTAAGAAAGCGAGGCTCTGAAGTCAGCTTCTTAGAAATCTACAGG	783
Db	1447	CCTAAAGACAGCTTTGTCTGTGGACGGGAAGAAATCACATTCCTGGAATACTACAGC	1506
Qy	784	AAGCAATACACCAAGAGATCACCGACTTGAAGCAGCGCTGTCTTGTTCAGCCAGCCCA--	841
Db	1507	AAAACCTATGGGATCA CAGTCAAGGAAGATGAC CAGCGCGTCTGTATCCACCGGCCAGT	1566
Qy	842	-AGAGAAGGGGGCCCTGGGGGACACTCCAGGGCTGCGCATCTCTTCTGCTGCCGAGCTC	900
Db	1567	GAGAGACAGAATAACCATGCGATCTTCTGAAGGGCGAGATCTCTGCTGCTGCCGAGCTC	1626
Qy	901	TGCTATCTTACAGTCTAACTGATAAATCGGTATATGATTTTAACTGATGAAGACTTAA	960
Db	1627	TCCTTCATCGGGGATCCCTCTGAAGATGAAGAAGGACTTCAGGGCCATGAAGGACTTG	1686
Qy	961	GCCGTTTCATACAGACTAACTCCAGAGCAAGGACGCTGAAGTGGGACGACTCAATTGAT	1020
Db	1687	ACTCAGCAGATTAACTGAGCCCCAAGCAGCACCAACGCTCTTTTGAATTCCTGTGCGAG	1746
Qy	1021	TACATTCATAAAGATATATGTTCAAAGGAGCTTCGACACTGGGGTTGAGCTTTGAT	1080
Db	1747	AGATTTTACAAAAACGAGGACGACGACAATGAGCTGACCGCTGGGGGCTCAGCTGCGAT	1806
Qy	1081	TCCAACCTTACTGTCTCTCTCAGGAAGAATTTGCAAAACAGAAAAAGATTCAACAGGTGA	1140
Db	1807	AAGATGTCCACAGATGAAGTTCGCTCTCTGCCATGGAGGATCACTTTAAGGAAC	1866
Qy	1141	AAAACTTTGATTAACAATCCAAATTTGCAGATTTGGTCCAAAGAAACAGAGGTGCACCA	1200
Db	1867	ACTTCAITTG---TCACATCGGAGGACCTGAACTGGGTTAAGGAAGTGACACAGATGCT	1923
Qy	1201	TTAATTAGTGTAAAGCCACTAGATAAAGTGGCTGTGATCTATACGGAAGAAATTTAGAA	1260
Db	1924	TCCAATCTAATATCCATGCAATTTCTGGGCACTCTTTTATCCAAAGAGACAAATGGAC	1983
Qy	1261	GCAGCCAAATCATGTATACAAAAATCTATTTAAAGTTACA CAGCCATGGGCATGCAAAATG	1320
Db	1984	CAAGCCAGAGAACTGGTTAACTGTTGAAAAGATTCGCCGGGCCAATGGCATCGGCATA	2043
Qy	1321	AGAAAAACAATAATGATTGGAAGT--GGAATGACAGAACTGAAGCCTACTTTAAGAGTCTTA	1377
Db	2044	AGCCCCCAGCCTGGTTGACCTGAAGGATGACCGAATAGAGACCTATATCAGGACCAIT	2103
Qy	1378	CAG-----CAAAAGGTCACAGCAGACACCCAGATATGTTGCTGTCTGTGTGTCAGATAT	1431
Db	2104	CAGTCTTACTTGGGAGTTGAGGGGAAGATACAAATGTTGCTGTTTGCATCATCATGGGCACA	2163
Qy	1432	CGGAAGGACAAATACGATGCTATTAAAAAAATACCTGTGTACAGATTGCGCCTACCCCAAGT	1491
Db	2164	CGTGATGATCTCTATGGAGCCATCAAGAGCTGTGCTGCTGCGTGAAGTCCCAAGTGCCTCA	2223
Qy	1492	CAGTGTGTGTGGGCCGAACCTTAGGCAAAACAGCAAACTGTCATGGGCCATTTGTCACAAAG	1551
Db	2224	CAGGTCATCAATGTCCGAACCATTTGTCAGCCCCCAGGCTTCGAGCGCTGGCTCAGAAA	2283
Qy	1552	ATTGCCCTACAGATGAATCTGAAGATGGGAGAGGCTCTGGAGGTGGACATCCCCCTG	1611
Db	2284	ATTTTACTTCAGATGAACCTGAACCTGGGTGTGAGCTCTGGGAGTGGATATTCGCTG	2343
Qy	1612	AAGCTCGTGATGATGTTGGGATCGATTTGATTTTACCATGATGACAGCTGGGCGGAGTCA	1671
Db	2344	AAACAACCTAATGGTGATTTGGATGGAATGTATACCATGACCCACAGAGGATGCGCTCT	2403
Qy	1672	ATCGCAGGATTTGTTGCCAGCATCAATGAAGGGATGACCCGCTGGTTCTCACTGCTGATA	1731

Db	2404	GTGGTCGGCTTCGTGGCCAGCATAAATCTCACACTACCAAATGGTACTCGAGGGTGGTG	2463
Qy	1732	TTTCAGGATAGAGGACAGGAGCTGTGTAGATGGGCTCAAAAGTCTGCCTGCAAGCGGCTCTG	1791
Db	2464	TTCCAGATGCCACATCAGGAGATTGTGGACAGCCTGAAGCTCTGCCTGTGGGTTCCCTTG	2523
Qy	1792	AGGGCTTGGAAATAGCTGCAATGAGTACATGCCAGCGCGGATCATCTGTGTACCCGATGGC	1851
Db	2524	AAAAAGTATTATCAGGTGAACCATTTCTCTCCACAGAGAAAATTTGTGTGTACCGAGATGGA	2583
Qy	1852	GTAGGAGAGCGCCAGCTGAAAACACGTGTGAACCTAGAAAGTGCACAGTTTTTTGGATTGT	1911
Db	2584	GTGCTGTATGGCCAGCTAAAGACAGTTGGCACTACAGATCCCTCAGCTGCAGAAAGTGT	2643
Qy	1912	CTAAAAATCCATTGGTAGAGGTTACAAACCTAGACTAACCGTAAATGTGTGTAAGAAAAAGA	1971
Db	2644	TTTGAAGCCTTTGATAAC--TACCACCCCAAGATGGTGTGTTGTAGTTCAAGAAGAAA	2700
Qy	1972	GTGAAACACAGATTTTTTGTCTAGTCTGTGAGGAAGACTTCAGNAATCCACTTCTCTGAACA	2031
Db	2701	ATCAGCACCAATCTGTACCTTGTGTCTCTGATCACTTCGTAAACCCCTCCCCCGGAGCT	2760
Qy	2032	GTATTATTGATGAGGTTTACCAGACAGAAATGGTAGACTTTTTTTATCGTAGACCCAGGCT	2091
Db	2761	GTGGTTGATCTATACCAATACAGACTGTGAGTGGGTGGATTTCTACCTTCTTGCCCATCAT	2820
Qy	2092	GTGAGAAGTGGTAGTGTTTTCTCCACACATTAACAATGTCACTATGACACACAGCGGCTG	2151
Db	2821	GTGCGACAGGCTGTGGCATACCTACACACTACATCTGTGTTCTGAAACATGCGAAATCTG	2880
Qy	2152	AAGCCAGACCAATACAGGCTTGTGACCTACAAGCTGTGCCACATCTATTACACTGGCCCA	2211
Db	2881	AGCCCTGATCACATGACAGGTTTGACTTTCAAACTATGCCACATGTACTGGAATTTGECCT	2940
Qy	2212	GGTGTCAATTCGTGTTCTGTCTCTTGCAGTAGCGCCCAAGCTGGCTTTTCTTGTGGCC	2271
Db	2941	GGTACCATCCAGATTCCAGTCTCTTGAAGTAGTGCCCAAGCTAGCTTCTCTGTCCGGA	3000
Qy	2272	CAGAGTATTCACAGAGAGCCAAATCTGTCACTGTCAAAACCGCCTTTACTACTCTATA	2328
Db	3001	CAGATTTTGATCATGAGCCAGCCATCCAGCTGTGTGGGAAGCTGTCTTCTTCGTATA	3057

RESULT 8
AA969938
LOCUS
DEFINITION
AA969938 462 bp mRNA linear EST 07-JUL-1998
OP60a11.s1 Soares NFL T GBC S1 Homo sapiens cDNA clone
IMAGE:1591212 3' Similar to WP:CO195.2 CE06748 ;, mRNA sequence.
ACCESSION
VERSION
KEYWORDS
SOURCE
ORGANISM
REFERENCE
AUTHORS
TITLE
JOURNAL
COMMENT
UNPUBLISHED
Contact: Robert Strausberg, Ph.D.
Email: cgapsb-@email.nih.gov
This clone is available royalty-free through LLNL ; contact the
IMAGE Consortium (info@image.llnl.gov) for further information.
Insert Length: 521 Std Error: 0.00
Seq primer: -40ml3 fwd. ET from Amersham.
Location/Qualifiers
1. .462
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="IMAGE:1591212"
/lab_host="DH10B"
FEATURES
source

BASE COUNT 127 a 97 c 124 g 114 t
 ORIGIN

Query Match 19.0%; Score 441.4; DB 9; Length 462;
 Best Local Similarity 99.6%; Pred. No. 9.4e-107;
 Matches 453; Conservative 0; Mismatches 1; Indels 1; Gaps 1;

1714 TGGTCTCAGCTGCATATTCAGGATAGAGCAGGAGCTGTAGATGGCTCAAGTC 1773
 Db 1 TGGTCTCAGCTGCATATTCAGGATAGAGCAGGAGCTGTAGATGGCTCAAGTC 60

1774 TGCCTGCAAGCGGCTCTGAGGGCTTGGATAGCTGCAATGAGTACATGCCAGCGGATC 1833
 Db 61 TGCCTGCAAGCGGCTCTGAGGGCTTGGATAGCTGCAATGAGTACATGCCAGCGGATC 120

1834 ATCGTGTACCGGATGGCTAGGAGCGGCCAGCTGGAACACTGGTGAACCTACGAAGTG 1893
 Db 121 ATCGTGTACCGGATGG-GTAGGAGCGGCCAGCTGGAACACTGGTGAACCTACGAAGTG 179

1894 CCACAGTTTTCGATGTCTAAATCCATTGTAGAGGTACACCCCTAGACTAAGGTA 1953
 Db 180 CCACAGTTTTCGATGTCTAAATCCATTGTAGAGGTACACCCCTAGACTAAGGTA 239

1954 ATTTGGTGAAGAAAGAGTGAACACCCAGATTTTTCGCTCAGCTGGAGGAGCTTCAG 2013
 Db 240 ATTTGGTGAAGAAAGAGTGAACACCCAGATTTTTCGCTCAGCTGGAGGAGCTTCAG 299

2014 AATCCATTCCTGGAACTGTTATGATGTAGAGTTTACCAGACCAAGATGGTATGACTTT 2073
 Db 300 AATCCATTCCTGGAACTGTTATGATGTAGAGTTTACCAGACCAAGATGGTATGACTTT 359

2074 TTTATCGTGAGCAGGCTGTGAGAGTGTAGTGTTCCTCCACACATTACATGTCATC 2133
 Db 360 TTTATCGTGAGCAGGCTGTGAGAGTGTAGTGTTCCTCCACACATTACATGTCATC 419

2134 TATGACACAGCGGCTGAAGCCAGCACATACA 2168
 Db 420 TATGACACAGCGGCTGAAGCCAGCACATACA 454

RESULT 9
 BG773137
 LOCUS
 DEFINITION 602721862F1 NIH_MGC_97 Homo sapiens cDNA clone IMAGE:4838918 5',
 mRNA sequence.
 ACCESSION BG773137
 VERSION BG773137.1 GI:14083790
 KEYWORDS EST.
 SOURCE Homo sapiens
 ORGANISM Homo sapiens (human)

REFERENCE
 1 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 NIH-MGC http://mhc.nci.nih.gov/
 National Institutes of Health, Mammalian Gene Collection (MGC)
 Unpublished
 Contact: Robert Strausberg, Ph.D.
 Email: cgapbs-remail.nih.gov
 Tissue Procurement: Miklos Palkovits, M.D., Ph.D.
 cDNA Library Preparation: Michael J. Brownstein (NHGRI), Shiraki

Toshiyuki and Piero Carninci (RIKEN)
 cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
 DNA Sequencing by: Incyte Genomics, Inc.
 Clone Distribution: MGC clone distribution information can be
 found through the I.M.A.G.E. Consortium/LLNL at:
<http://image.llnl.gov>
 High quality sequence stop: 789.
 Location/Qualifiers
 1. 818
 /organism="Homo sapiens"
 /mol_type="mRNA"
 /db_xref="taxon:9606"
 /clone="IMAGE:4838918"
 /lab_host="DH10B"
 /clone_lib="NIH MGC 97"
 /note="Organ: testis; Vector: pBluescriptR (modified
 pBluescript K3+); Site_1: BamHI; Site_2: SalI-XhoI (gtcgag
); Oligo-dr primed using primer 5'-TTTTTTTTTTT-3',
 size-selected for average insert size 2.2 kb and
 normalized to ROT 5. This is a primary library enriched
 for full-length clones and constructed using the
 Cap-trapper method (Carninci, in preparation). Library
 constructed by M. Brownstein (NIH/NHGRI, National
 Institutes of Health). Note: this is a NIH_MGC Library."
 BASE COUNT 245 a 174 c 198 g 201 t
 ORIGIN

Query Match 18.8%; Score 438.8; DB 12; Length 818;
 Best Local Similarity 97.3%; Pred. No. 6.4e-106;
 Matches 500; Conservative 0; Mismatches 7; Indels 7; Gaps 5;

8 TTGTGTGAACACAGGAGCAACTTAGACCATGTTAAAGATCAAAAACAGTCTTCAG 67
 Db 305 TTGTGTGAATACAGGAGCAACCTAGACCATGTTAAAGATCAAAAACAGTCTTCAG 364

68 GCATTATAGTAAAGTTAAGCACTAACCATTTCCGGCTGACATCCCGTCCCGAGTGGGCT 127
 Db 365 GCATTATAGTAAAGTTAAGCACTAACCATTTCCGGCTGACATCCCGTCCCGAGTGGGCT 424

128 TATATCATGTATCATATGACTATTAACCATCCATGATGGAGCCAGAGACTCCGTTCACTC 187
 Db 425 TATATCATGTATCATATGACTATTAACCATCCATGATGGAGCCAGAGACTCCGTTCACTC 484

188 TTCTTTTCAACACGAAAGATCTAATTGGAAGTGTCTATGCTTTTGTGGAACGATATAT 247
 Db 485 TTCTTTTCAACACGAAAGATCTAATTGGAAGTGTCTATGCTTTTGTGGAACGATATAT 544

248 TTTTACCTAAAGACTACAGAAAGGTTACTGAAGTTTGTAGTAAGACCCGGAATGGAG 307
 Db 545 TTTTACCTAAAGACTACAGAAAGGTTACTGAAGTTTGTAGTAAGACCCGGAATGGAG 604

308 AGGATGTGAGGATAACCATCATTTTACAAA-TGAACCTCCACCTACATACCACTTGT 366
 Db 605 AGGATGTGAGGATAACCATCATTTTACAAAATTGAACCTCCACCTACATACCACTTGT 664

367 TTGCAGTCTTATATATTTTTCAGGAGGCTTTTGAAGTGTCTATGCTTTTGTGGAACG 422
 Db 665 TTGCAGTCTTATATATTTTTCAGGAGGCTTTTGAAGTGTCTATGCTTTTGTGGAACG 724

423 AATTGGACGAAATTTATTAACCCAAATGACCAA-TTGATATTTCCAGTGCACAGTTGG 481
 Db 725 AATTGGACGAAATTTATTAACCCAAATGACCAAATTTGATATTTCCAGTGCACAGTTGG 784

482 TGATTTGG-CCTGGCTCACTACTTCCATCCTTC 514
 Db 785 TGATTTGGCCTGGCTCACTACTTCCATCCTTC 818

RESULT 10
 BU854142
 LOCUS
 DEFINITION BU854142 NIH_MGC_62 Homo sapiens cDNA clone IMAGE:6621314

363 DB AATGAGTACATGCCACCGAGATCATCTGTGTACCGGATGGCTAAGAAAGCCGACGTG 422

1870 QY AAAACACTGGT-GAACTACGAAAGT-GCCACAGTCTTTTGGATTGTCTAAAA 1917

423 DB AAAACACTGGTGGAACTACGAAAGTGGCCACAGTCTTTTGGATTGTCTAAAA 472

RESULT 11

BI:011107

LOCUS

DEFINITION

uvy19e03.yl McCarrey Eddy spermatoocytes Mus musculus cDNA clone

IMAGE:3658468 5' similar to TR:095404 095404 HIWI ;, mRNA sequence.

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

Mus musculus (house mouse)

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus

1 (bases 1 to 565)

Marra,M., Hillier,L., Kucaba,T., Martin,J., Beck,C., Wylie,T., Underwood,K., Steptoe,M., Theising,B., Allen,M., Bowers,Y., Person,B., Swaller,T., Gibbons,M., Pape,D., Harvey,N., Schurk,R., Ritter,E., Kohn,S., Shin,T., Jackson,Y., Cardenas,M., McCann,R., Waterston,R. and Wilson,R.

The WashU-NCI Mouse EST Project 1999

Unpublished

TITLE

JOURNAL

COMMENT

Contact: Marra M/WashU-NCI Mouse EST Project 1999

Washington University School of Medicine

4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA

Tel: 314 286 1800

Fax: 314 286 1810

Email: mouseest@wustl.edu

This clone is available royalty-free through LBNL ; contact the IMAGE Consortium (info@image.lbnl.gov) for further information.

MGI:1420772

Seq primer: Primer name ambiguous

High quality sequence stop: 418.

FEATURES

source

1..565

/organism="Mus musculus"

/mol_type="mRNA"

/strain="CD-1"

/db_xref="taxon:10090"

/clone="IMAGE:3658468"

/sex="male"

/tissue_type="spermatoocytes, pooled from multiple mice"

/dev_stage="60 day"

/lab_hosts="DH10B (phage-resistant)"

/clone_lib="McCarrey Eddy spermatoocytes"

/note="Organ: testis; Vector: pBluescript SK+ (Stratagene [5']- (GA)10-ACGAGTCGAGTGTCTTTT-3') and directionally cloned using 5' linkers 5'-AATCGGCACGAC-3' and 5'-CTCTGCGC-3'. Size selection of >400bp material gives average insert size ranging from 1-2 kb. Library was mass excised (from lambda-UniZAP-XR) and resulting single-stranded phagemids were prepped and transformed into DH10B. Library contains 98% recombinants. References: J. Androl. 20:635-639 and Gene 25:263-269. Library constructed and donated by J. McCarrey, Ph.D. (Southwest Foundation for Biomedical Research, Dept. of Genetics); excision done by E.M. Eddy, Ph.D. (National Institutes of Health, National Institute of Environmental Health Sciences). Original lambda-based library is available through ATCC, catalog #63422."

BASE COUNT 128 a 140 c 168 g 129 t

ORIGIN

Query Match 17.8%; Score 415.2; DB 10; Length 565;

Best Local Similarity 83.5%; Pred.No. 1.1e-99;

Matches 471; Conservative 0; Mismatches 93; Indels Gaps 0

```

QY 1677 AGGATTGTTCCAGCATCAATGAAGGATGACCCCTGTTCTCAAGTGCATATTTCA 1736
Db 1 AGGATTCGTCGCCAGCATCAATGAAGGATGACCCCTGTTCTCAAGTGCATATTTCA 60
QY 1737 GGATAGAGGACAGGAGCTGGTATGATGGCTCAAGTCTCCCTGCAAGCGCTCTGAGGC 1796
Db 61 GGACCGCGGAGGAGCTGGTATGATGGCTCAAGTCTCCCTGCAAGCGCTCTGAGGC 120
QY 1797 TTGGAATAGCTGCAATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1856
Db 121 TTGGAATAGCTGCAATGATGATGATGATGATGATGATGATGATGATGATGATGAT 180
QY 1857 AGACGCCAGCTGAAACACTGGTGAAGTACGAAGTGCACAGTTTGTGATGTTCTAAA 1916
Db 181 GGACGCGGAGCTGAAGACCTGGTCAATATGAGTCCACAGTTCCTAGATTGCTCAA 240
QY 1917 ATCCATTGTTAGAGGTTCAACCCCTAGACTTAACGGTAAATGTTGGTGAAGAAAGAGTGA 1976
Db 241 GTCACTCGGAGAGTTTACACCCAGACTGATGATGTTGGTGAAGAAAGAGTGA 300
QY 1977 CACCGAATTTTGTCTAGCTGAGGAGAGACTTCAGATCCACTTCCTGGAAACAGTTAT 2036
Db 301 TGCCAGGTTTTTGTCTAGCTGAGGAGAGACTTCAGATCCACTTCCTGGAAACAGTCA 360
QY 2037 TGATCTAGAGGTTACAGACCAAGATGATGATGATGATGATGATGATGATGATGAT 2096
Db 361 CGATGTGAAGTCAACAGACCAAGATGATGATGATGATGATGATGATGATGATGAT 420
QY 2097 AAGTGGTATGTTCTCCACACATTAACAATGATGATGATGATGATGATGATGATGAT 2156
Db 421 AAGCGGAGTGTGTTCCCAACACATTAACAATGATGATGATGATGATGATGATGAT 480
QY 2157 AGACCAATACAGCGCTTGACCTAAGCTGTGCAACTTATTAACAATGATGATGAT 2216
Db 481 CGACCAATACAGCGCTTGACCTAAGCTGTGCAACTTATTAACAATGATGATGAT 540
QY 2217 CATTCGTGTTCTGCTCCCTGCTGCA 2240
Db 541 CATCGAGTTCTGCACTGCTGCA 564

```

RESULT 12

```

AL041473/3
LOCUS DKEP434E2417 s1 434 (synonym: htes3) Homo sapiens cDNA clone
DEFINITION DKEP434E2417 3', mRNA sequence.
ACCESSION AL041473
VERSION AL041473.1 GI:5420824
KEYWORDS EST.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 413)
AUTHORS Bloecker,H., Boecker,M., Brandt,P., Mewes,H.W., Gassenhuber,J. and
Wiemann,S.
TITLE EST (Bloecker, et al.)
JOURNAL Unpublished
COMMENT Contact: Bloecker H
MIPS Ingolstraedter Landstr.1, D-85764 Neuherberg, Germany
This is the 3' sequence of the clone insert
Clone from S. Wiemann, Molecular Genome Analysis, German Cancer
Research Center (DKFZ); Email s.wiemann@dkfz-heidelberg.de;
sequenced by GBF (National Research Centre for Biotechnology Ltd.,
Braunschweig/Germany) within the cDNA sequencing consortium of the
German Genome Project.
ri sequence also available.
This clone (DKEP434E2417) is available at the RZPD in Berlin.
Please contact the RZPD: Ressourcenzentrum, Heubnerweg 6, 14059
Berlin-Charlottenburg, GERMANY; Email: clone@rzpd.de.
FEATURES Location/Qualifiers

```

source

```

1..413
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="DKEP434E2417"
/tissue_type="testis"
/dev_stage="adult"
/lab_host="DH10B"
/clone_lib="434 (synonym: htes3)"
/notice=vector: pSport1; Site_1: NotI; Site_2: SalI"
BASE COUNT 100 a 118 c 83 g 112 t
ORIGIN
Query Match 17.7%; Score 413; DB 9; Length 413;
Best Local Similarity 100.0%; Pred. No. 3.6e-99;
Matches 413; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1637 ATTGTTACCATGACATGACAGCTGGCGGAGGTCAATCGCAGGATTTGTTCCAGCATCA 1696
Db 413 ATTGTTACCATGACATGACAGCTGGCGGAGGTCAATCGCAGGATTTGTTCCAGCATCA 354
QY 1697 ATGAAGGATGACCGGCTGCTCAGCTGCATATTTTCAGATAGAGGACAGAGCTGG 1756
Db 353 ATGAAGGATGACCGGCTGCTCAGCTGCATATTTTCAGATAGAGGACAGAGCTGG 294
QY 1757 TAGATGGGCTCAAAAGTCTGCTGCAAGCGGCTCTGAGGGCTTGGATAGCTGCAATGAGT 1816
Db 293 TAGATGGGCTCAAAAGTCTGCTGCAAGCGGCTCTGAGGGCTTGGATAGCTGCAATGAGT 234
QY 1817 ACATGCCAGCGGATCATGCTGACCGGATGCGTAGGAGACGGCCAGCTGAAAAACAC 1876
Db 233 ACATGCCAGCGGATCATGCTGACCGGATGCGTAGGAGACGGCCAGCTGAAAAACAC 174
QY 1877 TGGTGAATGAGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGT 1936
Db 173 TGGTGAATGAGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGT 114
QY 1937 ACCCTAGACTAAACGGTAAATTTGGTGAAGAAAGAGTGAACACAGATTTTCTCAGT 1996
Db 113 ACCCTAGACTAAACGGTAAATTTGGTGAAGAAAGAGTGAACACAGATTTTCTCAGT 54
QY 1997 CTGAGGAAGACTTCAGATCCACTTCCTGGAACAGTATTGATGTAGAGTT 2049
Db 53 CTGAGGAAGACTTCAGATCCACTTCCTGGAACAGTATTGATGTAGAGTT 1

```

RESULT 13

```

BI560710
LOCUS 603254744F1 NIH_MGC_97 Homo sapiens cDNA clone IMAGE:5296954 5',
DEFINITION mRNA sequence.
ACCESSION BI560710
VERSION BI560710.1 GI:15448024
KEYWORDS EST.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 779)
AUTHORS NIH-MGC http://mgi.nci.nih.gov/.
JOURNAL National Institutes of Health, Mammalian Gene Collection (MGC)
COMMENT Unpublished
Contact: Robert Strausberg, Ph.D.
Email: cgabbs-r@mail.nih.gov
Tissue Procurement: Miklos Palkovits, M.D., Ph.D.
cDNA Library Preparation: Michael J. Brownstein (NHGRI), Shiraki
Toshiyuki and Piero Carninci (RIKEN)
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
Plate: LLAM11751 row: h column: 11

```

High quality sequence stop: 732.

FEATURES

source

Location/Qualifiers
 1. 779
 /organism="Homo sapiens"
 /mol_type="mRNA"
 /db_xref="taxon:9606"
 /clone="IMAGE:5296954"
 /lab_host="DH10B"
 /clone_lib="NIH_MGC_97"
 /note="Organ: testis; Vector: pBluescriptR (modified pBluescript KS+); Site 1: BamHI; Site 2: SalI-XhoI (Gtcgag); Oligo-dT primed using primer 5'-TTTTTTTTTTTTTTVN-3', size-selected for average insert size 2.2 kb and normalized to ROT 5. This is a primary library enriched for full-length clones and constructed using the Cap-trapper method (Carninci, in preparation). Library constructed by M. Brownstein (NIMH/NHGRI, National Institutes of Health). Note: this is a NIH_MGC Library."
 234 a 159 c 205 g 181 t

BASE COUNT

Query Match 17.4%; Score 405.6; DB 12; Length 779;
 Best Local Similarity 98.6%; Pred. No. 4.9e-97;
 Matches 430; Conservative 0; Mismatches 4; Indels 2; Gaps 2;

QY 8 TTGTTGTAACACAGGAGCACTTAGACCATGTTTAAAGAAATCAAAAACAGGTTCTTCAG 67
 DB 344 TTGTTGTAATACAGGAGCACTTAGACCATGTTTAAAGAAATCAAAAACAGGTTCTTCAG 403
 QY 68 GCATTATAGTAGTTAGGCTAATCACTTCCGGCTGACATCCCGTCCCGAGGGGCT 127
 DB 404 GCATTATAGTAGTTAGGCTAATCACTTCCGGCTGACATCCCGTCCCGAGGGGCT 463
 QY 128 TATATCATGATCACTATACCATACCACTGATGGAGCCAGAGACTCCGTTACGCTC 187
 DB 464 TATATCATGATCACTATACCATACCACTGATGGAGCCAGAGACTCCGTTACGCTC 523
 QY 188 TTCTTTTCAACAGAGATCAATTTGAAAGTGTGATGCTTTTGTGGAACGATATTAT 247
 DB 524 TTCTTTTCAACAGAGATCAATTTGAAAGTGTGATGCTTTTGTGGAACGATATTAT 583
 QY 248 TTTTACCTTAAAGACTACAGAAAGTTACTAGTTTGTAGTAAGACCGGAATGAG 307
 DB 584 TTTTACCTTAAAGACTACAGAAAGTTACTAGTTTGTAGTAAGACCGGAATGAG 643
 QY 308 AGGATGTGAGATAACGATCACTTTTAAAGATGATGCTTTTGTGGAACGATATTAT 367
 DB 644 AGGATGTGAGATAACGATCACTTTTAAAGATGATGCTTTTGTGGAACGATATTAT 702
 QY 368 TGCAGTTCTATATATTATTTTTCAGGAGGCTTTTGAATAATCA-TGAATTTGCAACAATT 426
 DB 703 TGCAGTTCTATATATTATTTTTCAGGAGGCTTTTGAATAATCA-TGAATTTGCAACAATT 762
 QY 427 GGCAGAAATTATTATA 442
 DB 763 GGCAGAAATTATTATA 778

RESULT 14

BU234973

LOCUS

DEFINITION
 603411470F1 CSEQCHN24 Gallus gallus linear EST 26-NOV-2002
 sequence.

ACCESSION

BU234973

VERSION

BU234973.1

KEYWORDS

EST

SOURCE

ORGANISM

Gallus gallus (chicken)

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Archosauria; Aves; Neognathae; Galliformes; Phasianidae;

Phasianinae; Gallus.

REFERENCE

1 (bases 1 to 664)

AUTHORS

Boardman, P.E., Sanz-Ezquerro, J., Overton, I.M., Burt, D.W., Bosch, E.,

TITLE
 JOURNAL
 MEDLINE
 PUBMED
 COMMENT

Fong, W.T., Tickle, C., Brown, W.R.A., Wilson, S.A. and Hubbard, S.J.
 A Comprehensive Collection of Chicken cDNAs
 Curr. Biol. 12 (22), 1965-1969 (2002)
 22335534
 12445392

Contact: Simon Hubbard

Department of Biomolecular Sciences

University of Manchester Institute of Science and Technology (UMIST)

PO Box 88, Manchester, M60 1QD, UK

Tel: 01612008930

Fax: 01612360409

Email: Simon.Hubbard@umist.ac.uk.

Location/Qualifiers
 1. 664

/organism="Gallus gallus"
 /mol_type="mRNA"
 /strain="White Leghorn, Hisex"
 /db_xref="taxon:9031"
 /clone="CHEST329K18"
 /dev_stage="22"
 /lab_host="DH10B"
 /clone_lib="CSEQCHN24"

/note="Organ: heads; Vector: pBluescript II KS(+); Site 1: EcoRI; Site 2: NotI; This normalized library was constructed from 1 million independent clones. cDNA synthesis was initiated using an oligo(dT) primer, using methylated C in the first strand synthesis reaction. Following this first strand reaction, double-stranded cDNA was blunt-ended, ligated to NotI adapters, digested with EcoRI, size-selected, and cloned into the NotI and EcoRI compatible sites of a custom modified MCS of the pBluescript (KS+) vector. The library was normalized in 2 rounds using conditions adapted from Soares et al., PNAS (1994) 91: 9228-9232 and Bonaldo et al., Genome Research 6 (1996): 791, except that a significantly longer reannealing hybridization was used."

BASE COUNT 182 a 132 c 162 g 188 t
 ORIGIN

Query Match 17.2%; Score 400.4; DB 13; Length 664;
 Best Local Similarity 76.5%; Pred. No. 1.1e-95;

Matches 491; Conservative 0; Mismatches 151; Indels 0; Gaps 0;

QY 1676 CAGGATTTGTTGCAGCATCAATGAGGATGATGACCCCTGGTCTTCTACGCTGCATATTC 1735
 DB 1 CTGGATTTGTTGGTAGCTGAATGAAAAATGACAGGTGGTTTTTTCACGCTGGCTTCTC 60
 QY 1736 AGGATAGAGCAGGAGCTGGTAGATGGCTCAAAGTCTGCTTGCCTGCAAGCGGCTGAGGG 1795
 DB 61 AAAGCCGTGGGAGGAAATTTGATGGGCTCAAAGCTGCTTGCAAACTGCTCTAAGGG 120
 QY 1796 CTTGGAATAGTCAATGATGATACATCCCGCGGATCATCGTGTATCCCGGATGGCGTAG 1855
 DB 121 AATGTTCAAGTGAATAAGTATTTGCCCTCTCGTATTAATTTGTTATCGTATGTTGTTAG 180
 QY 1856 GAGACCGCAGCTGAAAAACACTGGTGAACCTAAGTGGCACTGCTGTGATGCTGATGCTGTA 1915
 DB 181 GAGATGACAGCTCAATCTTTAGTGAACCTATGAAGTGGCTCAGTTTCTGATTTGTTGA 240
 QY 1916 AATCCATTGTTAGAGTTACACCCCTAGACTAACCGTAATTTGGTGAAGAAAGATGA 1975
 DB 241 AGACTGTTGTTAAAGACTCAATCCAAAGACTGACTGTGATCGTTGTTGAAGAAACGAGTGA 300
 QY 1976 ACACCAAGATTTTGTCTCAGTCTGGAGGAAGACTTCAGAAATCCACTTCTCTGGAACAGTTA 2035
 DB 301 GTACCAAGATTTCTTGGCAGGCTGGTGGAGGACTTAAACCACCCCTGTTGTTAGTCTGTG 360
 QY 2036 TTGATGTAGAGTTTACCAACAGAAATGTTATGACTTTTTTATCGTAGCCAGGCTGTGTA 2095
 DB 361 TTGATATAGAGGTGACCAACAGAAATGTTATGATTTCTTTTATTTGAGTCTAGGCAAGTGA 420
 QY 2096 GAAGTGGTAGTGTCTTCTCCACACATTCATCATGTCATCTATGACAAACAGCGGCTGAAGC 2155

403	GCATATAGTAAGGTTAAGCACTAAACCATTCGGGTGACATCCCGTCCCGAGTGGGCGT	462
128	TATATCAGTATCACATTGACTATAACCCCACTGATGGAAGCAGAACTCCGGTTCAGCTC	187
463	TATATCAGTATCACATTGACTATAACCCCACTGATGGAAGCAGAACTCCGGTTCAGCTC	522
188	TTCTTTTTCACACGCGAAGTCTTAATTTGGAAGAGTGTCATGCTTTTGTATGGAACGATATTAT	247
523	TTCTTTTTCACACGCGAAGTCTTAATTTGGAAGAGTGTCATGCTTTTGTATGGAACGATATTAT	582
248	TTTTCACATAAAGCACTACAGCAAAAGGTTACTGAACTTTTGTATGGAAGCAGAACTGGAG	307
583	TTTTCACATAAAGCACTACAGCAAAAGGTTACTGAACTTTTGTATGGAAGCAGAACTGGAG	642
308	AGAGTGTGAGGATAACGATCAGCTTTTAAACAAATGAACTTCCACTCATCACCMAACTTGT	367
643	AGAGTGTGAGGATAACGATCAGCTTTTAAACAAATGATCTTCAACTACAT--ACCAACTCTGT	700
368	TGCAGTTTCTATAATATTTTTCAGGAGGCTTTTGAATAATCATGAATTTGCAACAAATTG	427
701	TGAAAGTCTATATATTTTTCAGGAGGCG--TTTGAATAATCATGAATTTGAAACAAATTG	759
428	GACGAATTTATTTAACC---CAAATGACCCAAATTGATATTCAGTCCAGGTTGGTG--	483
760	GACGAATTTATTTAAGCACACAAATGAACCGAGTTGATATTCAAGTATAACGGTGGGG	819
484	-ATTGTGCGCTGGGTTCACTACTTCCCA-----TCCTTTCAGTATGAAAAACAGCATCATGCT	536
820	AATTGGACATGGATATACTACTTCCAAATCCCTTACAAATCTGATCAAAACGCAATCAGCT	879
537	CTGCAC--TGACGTTAGCCATAAGTCCCTCG--AGTGGAGACTGTTTGGATTTTCATGT	592
880	TTCTGCGGTGAGGTTTAGCCATAAAGTCGTTCGAAAGTGGAAGTGTGGGATTTTCAGGA	939
593	TC--AACTTTTATCATCAGACAGAGAACAATAAATTTCAAGAACAAAGTTTCCAAAGAACT	650
940	TCCAAATGTTTCATAATATAGCAGGAACAATTTATGTTTATGAACAGGTTTTCAGAGAAGT	999
651	AATAGGTTTATGTTGTTCTTACCAGGATATAACAAATAGACATACAGAGTGGATGATATCA	710
1000	-ATAGAATTATGTTGTTCTAGACAACATATACAATTAGACATCAAGAGGTAGTTATTATGA	1051
711	CTGGGACACAGAATCCCAAGAGCACCTTTTAAGAAA	744
1059	ATGCATCGAAATTAGATGATCTTTAGAAAGAAA	1092

Search completed: December 6, 2003, 15:49:50
Job time : 4864 secs

[illegible]

```

1059 ATGGCATCGAATTAGATGATCTTTAGAAAGAAA 1092

```